Tatouage robuste par étalement de spectre avec prise en compte de l'information adjacente

Gaëtan Le Guelvouit

N^o Ordre : D–03–15

THÈSE

présentée

DEVANT L'INSTITUT NATIONAL DES SCIENCES APPLIQUÉES DE RENNES

pour obtenir

le grade de : *DOCTEUR DE L'INSA DE RENNES*

Mention : *Informatique*

PAR

Gaëtan Le Guelvouit

Équipe d'accueil : TEMICS/IRISA/INRIA

École Doctorale : Mathématiques, Informatique, Signal, Électronique, Télécommunications (MATISSE)

TITRE DE LA THÈSE :

Tatouage robuste par étalement de spectre avec prise en compte de l'information adjacente

Soutenue le 24 novembre 2003 devant la commission d'Examen

COMPOSITION DU JURY

M.	CAMILLERAPP	Jean	Président
M.	DUHAMEL	Pierre	Rapporteurs
M.	PUN	Thierry	
M.	BAS	Patrick	Examinateurs
Mme	GUILLEMOT	Christine	
M.	NGUYEN	Éric	
M.	PATEUX	Stéphane	

À mes parents, à Jérôme

Remerciements

Ces travaux ont été effectués au sein du projet Temics de l'IRISA de Rennes. À ce titre, je remercie Christine Guillemot, directeur de recherche à l'INRIA, pour m'avoir accepté dans son équipe.

J'exprime tous mes remerciements à Jean Camillerapp, directeur du département Informatique de l'INSA de Rennes, qui m'a fait l'honneur de présider le jury. Je remercie Pierre Duhamel, directeur de recherche au CNRS, et Thierry Pun, professeur à l'Université de Genève, pour avoir accepté d'être les rapporteurs de cette thèse et pour leur remarques constructives sur ce manuscrit.

Je remercie également Patrick Bas, chargé de recherche au CNRS, pour avoir bien voulu juger ce travail. J'aimerais particulièrement remercier Éric Nguyen, à la fois pour son rôle en temps que membre du jury, mais aussi pour la collaboration que nous avons eue au sein du projet RNRT Diphonet, qui fût pour moi une grande source d'expérience. J'y associe Hervé Le Floch et Stéphane Baron de Canon Research France.

Merci à Stéphane Pateux pour avoir encadré ces travaux, pour son inépuisable patience, pour son regard critique et plus généralement pour tout ce que j'ai pu apprendre à son contact durant ce séjour. Ces quelques lignes ne peuvent exprimer toute ma gratitude.

Enfin, merci à tous les membres (anciens et nouveaux) du projet Temics qui ont rendu ces trois dernières années si agréables. Je salue particulièrement Luce Morin, Jérôme Viéron, Franck Galpin, Teddy Furon et Vivien Chappelier. Merci à Yoann, Pascal et Yacine pour leur soutien sans faille.

Table des matières

Introduction

L'ère numérique que nous traversons depuis quelques années a permis un accès à l'information bien plus aisé que par le passé. Les documents numériques étant immatériels, leur diffusion est extrêmement rapide et peu coûteuse. Les réseaux et les supports numériques de forte capacité facilitent les échanges de documents. Chacun peut copier, modifier et distribuer à son tour un fichier multimedia. Il est dans ce contexte très difficile de concilier libre accès à l'information et respect des auteurs et des droits associés. N'importe qui peut s'approprier une œuvre et l'utiliser en vue de faire des profits au dépend des ayant droits initiaux. Il est également difficile d'authentifier un document : on ne peut prouver qu'il n'a pas été retouché.

La cryptographie fut une première proposition pour sécuriser les documents numériques. Les schémas asymétriques (une clef secrète pour crypter et une clef publique différente pour décrypter) permettent de signer et donc d'authentifier un document. De plus, il est possible de contrôler la diffusion par la distribution des clefs : seuls les clients ayant acquitté des droits ont accès au document. Néanmoins, une fois décrypté, le document n'est plus protégé et peut être distribué malhonnêtement sans protection. Il est en outre impossible d'exposer librement les documents protégés. Par exemple, ce type de protection serait inapproprié pour une galerie en ligne.

Issu de la cryptographie et de la stéganographie[1], le tatouage (ou *watermarking*) numérique est une approche qui a fortement émergée depuis une dizaine d'années. Comme la stéganographie, le tatouage se propose de dissimuler au sein d'un document un message, tout en laissant le document marqué exploitable. Mais il y ajoute une notion de robustesse : la transmission du message doit être robuste aux modifications du media qui le véhicule.

Le tatouage trouve plusieurs applications dans la protection et l'authentification de document. Le tatouage **fragile** [LD99] permet de vérifier l'intégrité du document marqué. Il est très fragile aux modifications, et permet de vérifier que le document n'a pas été retouché et donc de l'authentifier (cas illustré par la figure 1(b)). Néanmoins, certains systèmes de tatouage fragile sont tout de même résistants aux traitements les plus usuels (compression avec perte notamment) afin de ne détecter que les modifications les plus préjudiciables vis-à-vis de l'interprétation du document. Ce type de

[1] L'art de dissimuler de l'information. Un des premiers exemples connus a été conté par Hérodote. Afin de transmettre un message, la tête d'un esclave fut tatouée. Après la repousse de ses cheveux, il a pu transmettre le message secret sans être inquiété [Sim98].

schéma de tatouage est dit **semi-fragile** [LC97, LBC+00]. Le message transmis est la plupart du temps un index issu d'une table de hachage appliquée sur le document à protéger.

À l'opposé, le tatouage **robuste** a pour objectif de transmettre une information malgré la modification du document. Lors de la lecture de la marque, certains algorithmes permettent d'extraire un message complet (une suite de symbole), tandis que d'autres indiquent simplement si le document a été marqué ou pas (on parle de **détection** de marque). Le tatouage robuste est particulièrement adapté au suivi et à la gestion de droits. Même si un fraudeur modifie le document, il est possible de retrouver l'auteur initial en insérant un numéro d'identification par tatouage robuste (illustré par la figure 1(c)). Le système *Image bridge* développé par Digimarc [Dig] est une application commerciale de ces techniques sur les images numériques. Les auteurs sont identifiés par un numéro et le tatouage de leurs images permet de retrouver les contrefaçons.

Outre ces différences dues à la robustesse, les schémas de tatouages peuvent être classés suivant les éléments nécessaires pour l'extraction (lecture du message depuis le document) de la marque. Un schéma **aveugle** n'a pas besoin du document original pour extraire. Au contraire, un schéma **non aveugle** nécessite le media original pour pouvoir lire correctement le message. Ces types de schémas sont de moins en moins étudiés, les applications concrètes étant assez rares.

Un dernier point discriminant est l'utilisation des clefs. La marque insérée est issue du codage du message à transmettre. Il est dépendant d'une clef. Si cette même clef est nécessaire au décodage (c'est-à-dire à l'extraction du message), le schéma est **symétrique** et dans le cas contraire, il est **asymétrique** (systèmes à clef privée et clef publique). On retrouve cette classification dans les algorithmes de cryptographie. De nombreuses recherches tentent de construire un schéma de tatouage asymétrique [FVD01, FD03, ESG00].

Cette étude se concentre sur le tatouage robuste et aveugle en vue de transmettre un message. De plus, nous nous plaçons dans le cadre d'un schéma symétrique. L'état de l'art nous propose actuellement deux alternatives : des schémas pratiques sous-optimaux, souvent basés sur des constatations empiriques, ou des schémas théoriques qui permettent d'exhiber des limites de performance mais dont la mise en œuvre pose problème. Notre objectif est de proposer un schéma de tatouage dont chaque choix est justifié, et pouvant être implémenté sans difficulté. Afin d'optimiser notre technique, nous utiliserons la théorie des jeux. Nous proposerons également une technique de codage adaptée aux canaux avec information adjacente (modèle théorique convenant particulièrement au tatouage). Nos résultats nous permettrons de justifier certains choix empiriques déjà proposés par la communauté scientifique, mais aussi d'introduire des techniques inédites.

Ces travaux ont été menés dans le cadre du projet RNRT Diphonet. Son objectif est de proposer une plate-forme d'authentification et de suivi d'images à destination des

professionnels de la photographie (agences de presses, photographes indépendants, ...). Grâce au tatouage robuste, les utilisateurs pourront protéger leurs œuvres, et suivre leur diffusion sur l'Internet par un système d'exploration automatique du réseau.

Plan du manuscrit

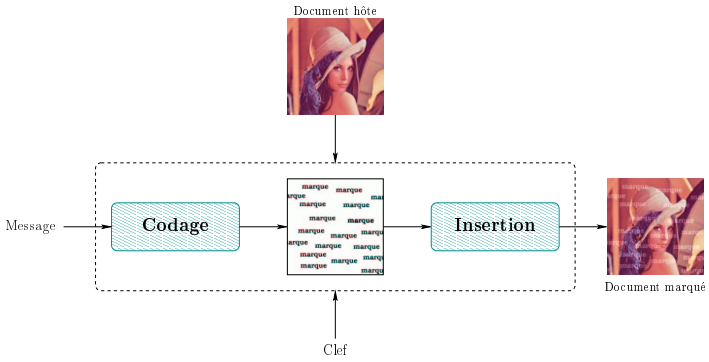
La première partie du manuscrit porte sur l'étude de l'existant. Nous étudierons les méthodes de tatouage les plus représentatives de l'état de l'art, mais aussi les domaines associés à cette problématique. Ainsi, nous verrons quelques exemples de modèles perceptuels utilisés afin de limiter la perception de la marque ou encore les principaux résultats en codage canal utilisés par la suite.

S'appuyant sur une transmission par étalement de spectre, la seconde partie développe l'optimisation de notre schéma de tatouage. L'interaction entre la phase d'insertion de la marque et la transmission du document marqué est vue comme un jeu entre un attaquant et un défenseur. La théorie des jeux nous permet alors de définir une forme d'attaque optimale, puis la défense adaptée. Les expérimentations appliquées sur des images montrent l'intérêt d'une telle approche.

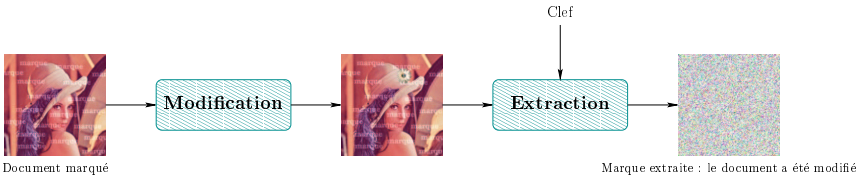
En tirant profit des canaux avec information adjacente, les travaux de Costa [Cos83] ont montré un gain théorique très important. Dans la troisième partie, nous proposons tout d'abord une implémentation pratique du schéma de Costa, grâce à la construction de codes adaptés. Nous introduisons de plus un algorithme itératif simple permettant de supprimer l'interférence inter-symboles, point faible de l'étalement de spectre. Nous étendons ensuite notre jeu en introduisant la prise en compte de désynchronisations géométriques, puis grâce à une modélisation statistique plus précise, nous définissons une attaque **informée**. Enfin, le dernier chapitre présente une application pratique de nos résultats. Nous définissons un schéma de tatouage d'images utilisant la transformée en ondelettes. Grâce à nos résultats précédents, nous montrons comment choisir une stratégie d'insertion permettant de garantir un niveau de robustesse et comment appliquer nos résultats à des scénarii d'utilisation.

Notations

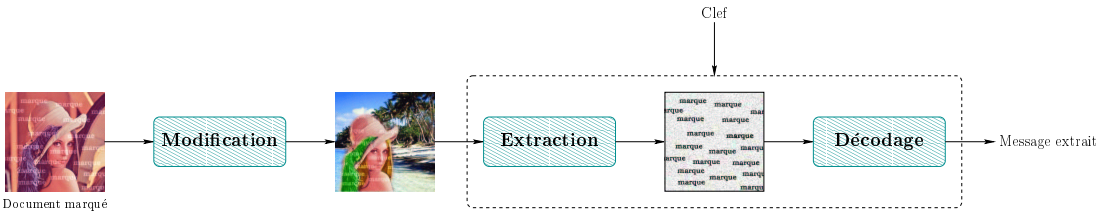
Les notations suivantes seront utilisées par la suite :
- $\mathsf{x} = \{x_1, x_2, \ldots, x_n\}$: vecteur,
- $\mathcal{N}(\mu, \sigma^2)$: loi Normale de moyenne μ et de variance σ^2,
- $\langle \mathsf{x}, \mathsf{y} \rangle$: produit scalaire entre les vecteurs x et y,
- $x \propto y$: équivalement à $x = k \times y$ avec k constante indépendante de x et de y,
- $\mathbf{P}_{\mathrm{e}}^{\mathrm{b}}$: probabilité d'erreur par bit,
- $\mathsf{X} = \{X_1, X_2, \ldots, X_n\}$: ensemble de variables aléatoires, dont la réalisation est le vecteur x,
- $\mathbb{E}[X]$: espérance de la variable aléatoire X,
- $H(X)$: entropie de la v.a. X, donnée par $-\sum_x \Pr(X = x) \times \log_2 \Pr(X = x)$,
- $I(X;Y)$: information mutuelle entre les v.a. X et Y, donnée par $H(X) - H(X|Y)$.



(a) Tatouage d'une image



(b) Tatouage fragile : la moindre modification du document se répercute fortement sur la marque extraite et on peut en déduire que le document n'est pas authentique



(c) Tatouage robuste : malgré de fortes modifications, le message doit pouvoir être lu correctement

FIG. 1 – Différentes utilisations du tatouage : fragile ou robuste

De plus, nous nous efforcerons de conserver les mêmes notations pour les différents éléments impliqué dans le tatouage :

- $\mathsf{m} = \{m_1, m_2, \ldots, m_k\} \in \mathcal{M}^k$: message (suite de k symboles) transmis par le tatouage,
- $\mathsf{x} = \{x_1, x_2, \ldots, x_n\} \in \mathcal{X}^n$: signal hôte issu du document à tatouer,
- $\mathsf{w} \in \mathcal{X}^n$: marque, obtenue par le codage d'un message,
- $\mathsf{y} \in \mathcal{X}^n$: signal tatoué (ou marqué),
- $\mathsf{y}' \in \mathcal{X}^n$: signal marqué puis attaqué,
- $c \in \mathcal{C}$: clef utilisée pour le codage ou/et le décodage,
- D_{xy} : mesure la distorsion entre x et y, c'est-à-dire la distorsion introduite par le tatouage,
- D_{a} : distorsion introduite par les attaques. Cette mesure peut correspondre à $D_{\mathsf{yy}'}$ ou $D_{\mathsf{xy}'}$.

Première partie

Contexte général

Chapitre 1

Le tatouage aveugle robuste

Le tatouage, dans la version qui nous intéresse dans ce manuscrit, consiste à transmettre un message *via* un document hôte, c'est-à-dire en utilisant ce document comme support. Ce dernier est donc modifié de façon à ce qu'un message puisse en être extrait après analyse, sans que son exploitation normale soit remise en cause : la modification ne doit pas être perceptible au point de gêner l'utilisation du document tatoué. Cette étude se focalise sur la version aveugle et robuste du tatouage. Un schéma de tatouage est dit **aveugle** si l'on peut extraire le message uniquement à partir du document marqué, et donc sans le document original. Un schéma est **robuste** si, malgré le fait que le document marqué ait été modifié, il reste possible d'extraire la marque sans erreur. La robustesse d'un schéma se quantifie par le nombre ou la force des modifications que peut subir le document marqué sans mettre à mal l'extraction du message. Le terme robuste est principalement utilisé pour se démarquer du tatouage fragile, où toute modification influe sur l'extraction de la marque (dans le but de vérifier l'intégrité du document) et ne correspond pas à une norme quelconque de robustesse. Seuls les schémas de tatouage symétriques seront étudiés. Les phases d'insertion et d'extraction sont paramétrées par une clef unique (clef privée). Il existe néanmoins plusieurs schémas proposant une approche par clef privée et publique [FVD01, FD03, ESG00].

1.1 Transmission d'un message par le tatouage d'un document

Transmettre un message par le support d'un document comporte trois phases : l'insertion du message dans le support, la diffusion du document marqué et enfin l'extraction du message. Cet enchaînement est résumé par la figure 1.1, dont les briques sont détaillées dans les sections suivantes.

1.1.1 Insertion du message

La totalité des schémas de tatouage proposés par la communauté scientifique est modélisable par l'enchaînement d'actions décrit par la figure 1.2. Les notations des

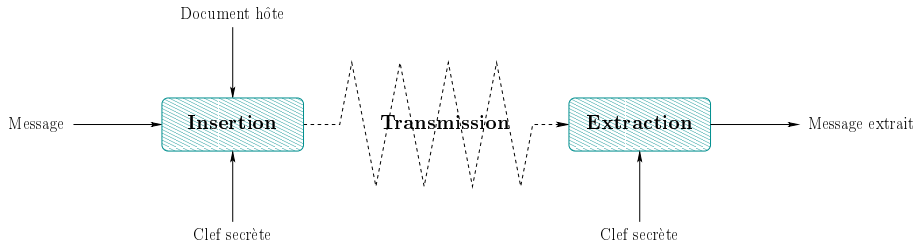


FIG. 1.1 – Transmission d'un message par le tatouage d'un document

différents éléments de cette figure seront utilisées dans le reste de ce document.

Les données hôtes, modifiées par le processus d'insertion, sont notées sous la forme d'un vecteur d'éléments $\mathsf{x} = \{x_1, x_2, \ldots, x_n\} \in \mathcal{X}^n$. Elles sont obtenues à partir du document d'origine. Par exemple, si le document hôte est une image monochrome, chaque élément x_i correspond à la luminosité du $i^{\text{ème}}$ pixel de l'image, avec $\mathcal{X} = \{0, 1, \ldots, 255\}$. Néanmoins, il peut être intéressant de changer l'espace de représentation du document hôte afin d'obtenir un signal x aux meilleures propriétés (exploitation d'un modèle perceptuel, meilleur comportement face à certains types de modifications, ...). Ce changement est indiqué par l'étape de transformation de la figure 1.2 et est étudié plus en détails dans la section 2.

Comme dit dans l'introduction, nous nous penchons sur la transmission d'un message, et non sur la détection d'une marque[1]. Un message est une suite de k symboles $\mathsf{m} = \{m_1, m_2, \ldots, m_k\} \in \mathcal{M}^k$. Dans le cas de messages binaires, nous avons donc $\mathcal{M} = \{0, 1\}$. Le message est codé (codes correcteurs d'erreur, modulation, ...) avant insertion, donnant le signal de la marque noté $\mathsf{w} = \{w_1, w_2, \ldots, w_n\} \in \mathcal{X}^n$. Ce codage peut se faire en prenant en compte le signal hôte (pour adapter l'énergie de la marque en fonction du signal ou pour prendre en compte une information adjacente, comme présenté dans la section 3.3). La forme du codage est dépendante d'une clef $c \in \mathcal{C}$, nécessaire au décodage, disponible uniquement lors de la phase d'insertion et d'extraction. Le signal marqué est alors donné[2] par $\mathsf{y} = \mathsf{x} + \mathsf{w}$. Enfin, la transformée inverse de celle utilisée en début de schéma (changement de représentation) termine cette phase d'insertion. L'ajout d'un signal de marque au signal hôte introduit une distorsion inévitable entre le document original et sa version tatouée. Elle est notée D_{xy}. On considère donc l'insertion d'une marque comme une fonction

$$\begin{array}{rcl} \mathcal{X}^n \times \mathcal{M}^k \times \mathcal{C} \times \mathbb{R} & \longrightarrow & \mathcal{X}^n \\ \text{insert} : (\mathsf{x}, \mathsf{m}, c, D_{\mathsf{xy}}) & \longmapsto & \mathsf{y}. \end{array} \tag{1.1}$$

[1] Extraire un message ou détecter une marque comporte néanmoins de nombreuses similarités. Cela est discuté dans la section 1.3

[2] Le tatouage substitutif peut aussi être vu comme l'ajout de deux signaux, le signal de marque étant adapté au signal hôte, d'où l'intérêt du lien entre signal hôte et codage

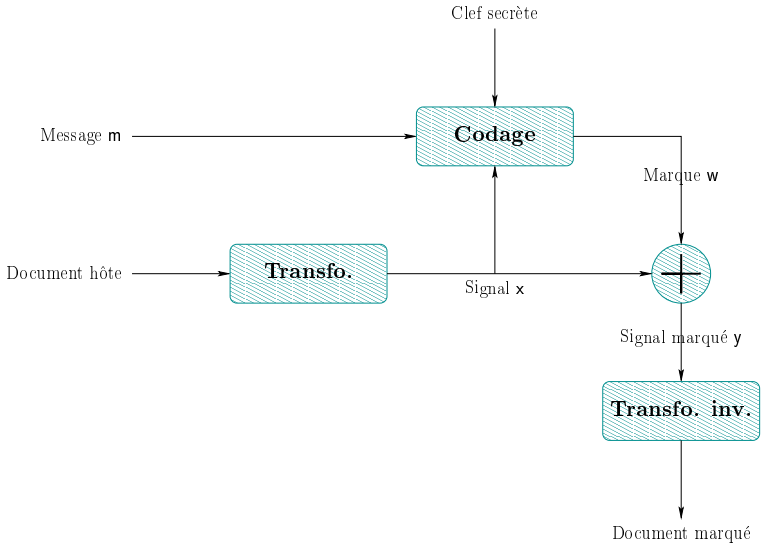


FIG. 1.2 – Schéma d'insertion classique d'une marque au sein d'un document

1.1.2 Diffusion et attaques

Entre le marquage et l'extraction de la marque, le document est diffusé. Cela s'accompagne d'éventuelles modifications : il est parfois nécessaire de changer de format, de passer par des canaux analogiques (avec les conversions adéquates), ... On distingue les attaques involontaires des attaques malicieuses. Dans le premier cas, le document est modifié par commodité : on y classe les compressions avec perte (passage au format JPEG d'une image, compression de musique en MP3), les conversions A/N et N/A ou encore les traitements de document usuels (filtrage, égalisation, changement d'échelle, ...). Les attaques malicieuses ont pour but d'atténuer la marque (l'optimal étant de la supprimer totalement) afin de rendre l'extraction du message initialement inséré impossible. De nombreux articles ont proposé de telles attaques [VDP00, HVR01, SEG01b, RDCD02], adaptées à un type de tatouage précis. On considère que l'attaque, volontaire ou non, est indépendante de la clef utilisée lors de l'insertion : même si l'attaquant connaît parfaitement la technique de tatouage utilisée, l'utilisation d'une clef secrète fait qu'il ne peut connaître la forme exacte du signal w ajouté, comme le préconise le principe de Kerckoffs [Ker83].

Hormis quelques exceptions (par exemple les attaques sur le protocole de transmission de la clef secrète, non étudiées ici), les attaques peuvent se classer en deux catégories, aux conséquences très différentes vis-à-vis du tatouage. Les attaques de type traitement du signal vont modifier individuellement les échantillons de y : le $i^{\text{ème}}$ échantillon y'_i reçu correspondra bien au $i^{\text{ème}}$ échantillon marqué, avec les effets de l'attaque en plus (filtrage, bruit supplémentaire, ...). Les attaques désynchronisantes ont pour but de détruire cette correspondance. Les échantillons sont décalés, et le message ne pourra pas être extrait correctement. Il est indispensable d'appliquer un

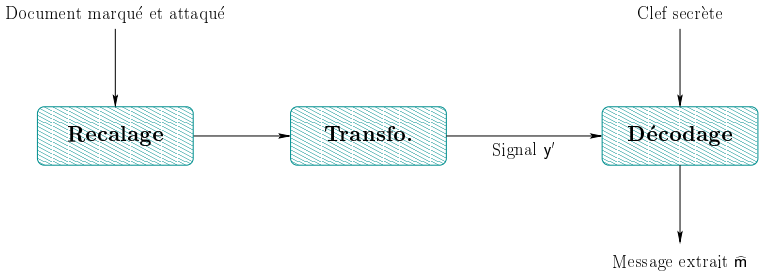


FIG. 1.3 – Schéma d'extraction classique d'une marque depuis un document

module de recalage avant l'extraction. Les modifications géométriques (rotations, translations, ...) sont des attaques désynchronisantes.

L'attaque est vue comme une fonction dépendante de la distorsion introduite, c'est-à-dire

$$\begin{aligned} \mathcal{X}^n \times \mathbb{R}^+ &\longrightarrow \mathcal{X}^n \\ \text{attaque} : (\mathsf{y}, D_{\mathrm{a}}) &\longmapsto \mathsf{y}'. \end{aligned} \tag{1.2}$$

Il y a deux manières de mesurer D_{a} : soit considérer $D_{\mathsf{yy}'}$ et donc la distorsion entre le document marqué et sa version attaquée, soit prendre $D_{\mathsf{xy}'}$, c'est-à-dire la distorsion entre le document original et sa version marquée puis attaquée. Il est entendu que cette dernière ne peut être mesurée directement par l'attaquant, celui-ci n'ayant évidemment pas accès aux données hôtes d'origine. Néanmoins, cette version comporte le principal avantage de justifier l'emploi de filtres débruiteurs en guise d'attaque [MI01b].

1.1.3 Extraction du message

À moins d'utiliser une transformation invariante aux attaques désynchronisantes, il est indispensable de disposer d'un module de recalage, appliqué avant la transformation et le décodage du signal reçu. La phase d'extraction reprend les étapes équivalentes à celles de l'insertion : changement de représentation aboutissant au signal marqué et attaqué noté y', puis décodage de ce signal (figure 1.3). Cet enchaînement définit

$$\begin{aligned} \mathcal{X}^n \times \mathcal{C} &\longrightarrow \mathcal{M}^k \\ \text{extraction} : \big(\mathsf{y}', c\big) &\longmapsto \widehat{\mathsf{m}}. \end{aligned} \tag{1.3}$$

1.1.4 Mesures de performance

Il est indispensable de pouvoir quantifier la performance d'un schéma de tatouage. Elle peut varier en fonction du document hôte à marquer (certains signaux sont facilement attaqués et le schéma de tatouage sera moins performant), de la taille du

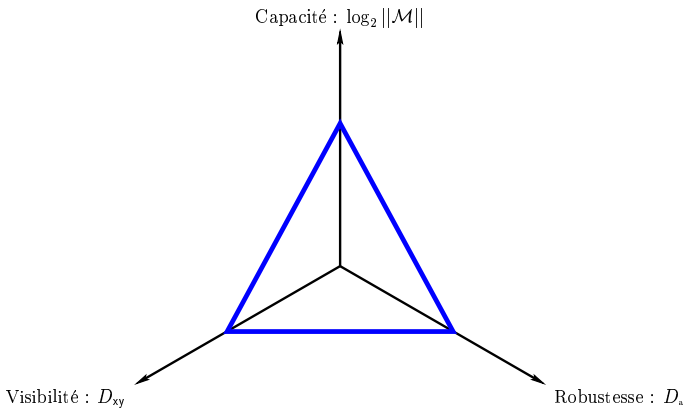


FIG. 1.4 – Représentation schématique du compromis entre robustesse, capacité et visibilité

message transmis, et surtout des distorsions d'insertion et d'attaque. La performance d'un schéma de tatouage est donc une fonction

$$\begin{array}{rcl} \mathcal{X}^n \times \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^+ & \longrightarrow & \mathbb{R} \\ \mathrm{perf} : \left(\mathsf{x}, \log_2 \|\mathcal{M}^k\|, D_{\mathsf{xy}}, D_{\mathrm{a}}\right) & \longmapsto & p. \end{array} \tag{1.4}$$

Une mesure intéressante est la probabilité d'erreur d'extraction, notée $\mathbf{P}_{\mathrm{e}}$: on mesure la probabilité que le message extrait $\widehat{\mathsf{m}}$ soit différent du message inséré. Une autre possibilité est de mesurer la capacité du canal formé par le tatouage, c'est-à-dire la taille maximale de message qu'il est théoriquement possible de transmettre sans erreur. Même si la capacité permet effectivement de quantifier la performance d'un schéma, elle n'a pas de répercussion pratique directe. Dans le cadre du tatouage robuste, on cherche à maximiser la robustesse du schéma, et non à transmettre le plus grand message possible. La taille du message est généralement fixée et est commune à l'insertion et l'extraction. Enfin, une mesure complémentaire est la probabilité de fausse alarme, notée $\mathbf{P}_{\mathrm{f}}$: elle mesure la probabilité d'extraire un message alors que le document n'a pas été tatoué. Elle est surtout utilisée dans le cas de détection de marque, où les fausses alarmes réduisent la confiance et donc l'intérêt du tatouage.

Quel que soit le schéma de tatouage, si l'on souhaite atteindre une performance donnée, il faut faire un compromis entre la distorsion d'insertion, la taille du message et la distorsion d'attaque que le schéma va pouvoir supporter. Cela est illustré par la figure 1.4. Par exemple, mettre des messages de taille plus importante sans modifier D_{xy} induit que le schéma sera moins robuste.

1.2 Mises en pratique

La construction du signal de marque w lors de l'insertion et son extraction depuis le document reçu caractérise les techniques de tatouage. On classe ces techniques en

deux catégories : d'abord les techniques additives, où le signal ajouté n'est pas corrélé au signal hôte, puis les techniques substitutives, où les données hôtes sont modifiées afin de correspondre à un message codé.

1.2.1 Tatouage additif

Le tatouage additif consiste à ajouter un signal w à x ($\mathsf{y} = \mathsf{x} + \mathsf{w}$), sans que le codage amenant à w soit déterminé par x, même si les échantillons w_i peuvent être modulés par un facteur perceptuel dépendant de x_i (voir la section 4). Classiquement, on pose $\mathbb{E}\left[\langle \mathsf{x}, \mathsf{w} \rangle\right] = 0$. L'extraction se fait en décodant le signal y' reçu, c'est-à-dire en décodant w bruité par l'attaque et par le signal hôte x (voir la figure 1.5). De plus, afin de respecter la contrainte d'imperceptibilité, l'énergie de w est très inférieure à celle de x. Nous sommes en face d'un système de transmission très fortement bruité.

La fonction de codage permet d'associer à un message $\mathsf{m} \in \mathcal{M}^k$ un signal $\mathsf{w} \in \mathcal{W}$ avec $\mathcal{W}^n \subset \mathcal{X}^n$: codage$(\mathsf{m}, c) = \mathsf{w}$. De même, la fonction de décodage permet de retrouver un message depuis un signal reçu, c'est-à-dire le message correspondant au signal $\widehat{\mathsf{w}}$ le plus proche du signal reçu. Dans le cas où $\|\mathcal{M}^k\|$ est important, la recherche du signal le plus proche est en pratique impossible (pour un message binaire de n bits, il faudrait 2^n calculs de distance). L'étalement de spectre [PWM82] est une solution de communication adaptée à ce type de caractéristiques et utilisable sans difficulté pour des valeurs de $\|\mathcal{M}^k\|$ importantes. Le principe est de coder les symboles de m séparément, plutôt que de coder le message complet. Soit la fonction d'étalement

$$\begin{aligned} \{0, 1, \ldots, k\} \times \mathcal{M} \times \mathcal{C} &\longrightarrow \mathcal{X}^n \\ \mathrm{ss} : (j, m, c) &\longmapsto \mathsf{w}_j. \end{aligned} \tag{1.5}$$

Chaque symbole est codé différemment selon son rang dans m ($\mathrm{ss}(i, m_i, c) \neq \mathrm{ss}(j \neq i, m_i, c)$). Le vecteur de marque est donné par

$$\mathsf{w} = \sum_{j=1}^{k} \mathrm{ss}(j, m_j, c). \tag{1.6}$$

L'extraction du $j^{\text{ème}}$ symbole consiste à rechercher le signal $\widehat{\mathsf{w}}_j$ le plus proche de y' parmi les $\|\mathcal{M}\|$ possibles. Pour un message de n symboles binaires, l'extraction complète se fait en seulement $2n$ calculs de distance.

Dans le cas de symboles binaires, une simplification commune [CDRP99a, CDRP99b] consiste à prendre $\mathcal{M} = \{-1, +1\}$. La fonction ss() est définie par une matrice $\mathsf{G} \in \mathcal{X}^{k.n}$ de dimensions $k \times n$ générée pseudo-aléatoirement depuis la clef c. Les w_j sont obtenus grâce à la modulation du $j^{\text{ème}}$ vecteur de G (c'est-à-dire la $j^{\text{ème}}$ porteuse) par le $j^{\text{ème}}$ bit :

$$\begin{aligned} \{0, 1, \ldots, k\} \times \{-1, +1\} \times \mathcal{C} &\longrightarrow \mathcal{W}^n \\ \mathrm{ss} : (j, m, c) &\longmapsto m_j \times \mathsf{G}(j). \end{aligned} \tag{1.7}$$

Pour des signaux gaussiens, le calcul de la distance entre y' et une porteuse consiste en un calcul de produit de corrélation entre ces deux signaux :

$$\widehat{m}_j = \arg\max_{m\in\mathcal{M}} \left\{ \langle \mathsf{y}', \mathrm{ss}(j,m,c)\rangle \right\}. \tag{1.8}$$

Si l'on prend $\mathsf{y}' = \mathsf{y} = \mathsf{x} + \sum_{j=1}^{k} \mathrm{ss}(j, m_j, c)$ (signal marqué non attaqué), le résultat de ce calcul est séparable en trois composantes :

$$\begin{aligned} \langle \mathsf{y}', \mathrm{ss}(j,m_j,c)\rangle &= \langle \mathrm{ss}(j,m_j,c), \mathrm{ss}(j,m_j,c)\rangle + \langle \mathsf{x}, \mathrm{ss}(j,m_j,c)\rangle \\ &\quad + \sum_{l=1,l\neq j}^{k} \langle \mathrm{ss}(j,m_l,c), \mathrm{ss}(j,m_j,c)\rangle. \end{aligned} \tag{1.9}$$

On trouve donc :

- l'auto-corrélation de la porteuse. C'est dans ce terme d'espérance non nulle que réside d'énergie de la marque,
- le produit de corrélation entre la porteuse et le signal hôte. Malgré son espérance nulle, ce terme est la principale source d'interférence (et donc de potentielles erreurs),
- le produit de corrélation entre la porteuse du $j^{\text{ème}}$ symbole et celle des autres symboles. Bien que dans l'idéal les porteuses doivent être parfaitement orthogonales afin de s'assurer d'une interférence nulle, leur grand nombre et leur grande longueur rend cette construction quasi impossible. Cela entraîne une interférence résiduelle (interférence inter-symboles, appelée aussi ISI) limitant la performance du canal de transmission.

La transmission par étalement de spectre définit un canal gaussien [PBBC97, PBBC98] que l'on caractérise par son rapport signal-à-bruit E_b/N_0. Ce type de canal est détaillé dans la section 3.1.2.

Outre le classique calcul de corrélation de l'équation 1.8, Cox *et al.* [CMB02] proposent de normaliser la mesure en la divisant par $|\mathsf{y}'|.|\mathrm{ss}(j,m,c)|$. Dans la même optique, il a été proposé de prendre en compte l'attaque qu'a subit le signal marqué en divisant le produit de corrélation par la somme des énergies de la marque et du bruit d'attaque ajouté [WYL02]. Une autre possibilité est, plutôt que d'utiliser y' dans le calcul, de prendre une version débruité du signal reçu (en vue d'atténuer le bruit introduit par le signal hôte). Voloshynovskiy *et al.* [VDPP01] utilisent pour cela un filtre de Wiener.

1.2.2 Tatouage substitutif

Plutôt que de construire un signal w n'ayant que peu de rapport avec les données hôtes, le tatouage substitutif se propose de modifier ces données afin de les faire correspondre au message que l'on souhaite transmettre. On peut classer dans ces méthodes le tatouage par quantification, vu en détail dans la suite de cette section, et les techniques de tatouage imposant un ensemble de contraintes aux données marquées[3].

[3] Tatouage dit **virtuel** [Man01].

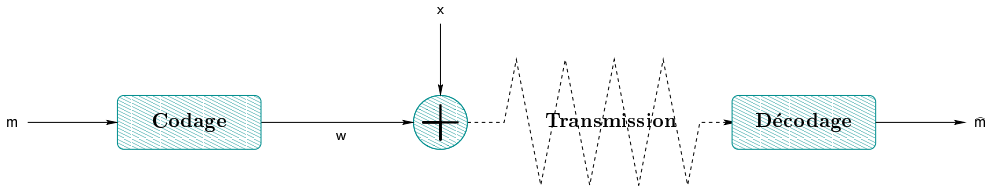


FIG. 1.5 – Codage et décodage d'un message transmis par tatouage additif

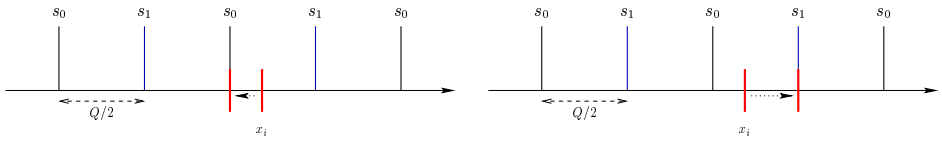
Quantification

Une façon très simple de tatouer une image est d'utiliser les bits de poids faible (LSB) des échantillons de x [Tur89] : ils peuvent être modifiés sans grand impact perceptuel. Ces bits sont alors forcés aux valeurs correspondant au message m. Cette simplicité se paye par une très faible robustesse : n'importe quel traitement, même peu important, suffit à modifier les LSB et donc à rendre l'extraction impossible. Cette technique est à classer dans le tatouage substitutif (la donnée x_i est remplacée par une donnée y_i très proche) et plus précisément dans les méthodes de tatouage par quantification.

Le principe du tatouage par quantification est de substituer les données hôtes par des états de quantification. L'extraction se fait en prenant l'état le plus proche des données reçues. Considérons un tatouage par quantification m-aire. Les données à marquer seront donc modifiées par sous-signaux de dimension m. Le nombre de symboles à insérer n'étant pas forcément égal à n/m, le message est d'abord codé par la fonction codage$(\mathsf{m}, c) = \mathsf{v}$ tel que $\mathsf{v} \in \mathcal{M}^{n/m}$. Chaque symbole ainsi obtenu sera inséré dans un des n/m sous-signaux. Soit un ensemble $\{Q_{s_1}, Q_{s_2}, \ldots, Q_{s_{\|\mathcal{M}\|}}\}$ de quantificateurs m-aires associés aux $\|\mathcal{M}\|$ symboles possibles. Pour chaque symbole v_j de v, le quantificateur Q_{v_j} définit un ensemble d'états de quantification possibles. L'état le plus proche du sous-signal considéré est le sous-signal marqué.

L'extraction du $j^{\text{ème}}$ symbole de $\widehat{\mathsf{v}}$ se fait prenant tous les états de quantification possibles avec les $\|\mathcal{M}\|$ quantificateurs. Le quantificateur dont un des états est le plus proche du sous-signal reçu donne le symbole extrait $\widehat{v}_j$ auquel il est associé. La suite de symboles $\widehat{\mathsf{v}}$ est décodée par l'inverse de la fonction codage() pour donner $\widehat{\mathsf{m}}$. Le principe est illustré par la figure 1.6 dans le cas d'une quantification scalaire ($m = 1$) et de symboles binaires ($\|\mathcal{M}\| = 2$).

L'avantage de cette méthode est qu'il est possible de définir un ensemble de quantificateurs tel qu'il n'y ait aucune erreur pour un niveau de bruit donné (contrairement à l'étalement de spectre vu au-dessus, à cause des interférences lors de la corrélation d'extraction). Ainsi, si le bruit ajouté est inférieur à la distance minimale entre deux états de quantification ($Q/2$ dans le cas montré par la figure 1.6), l'extraction est sans erreur. De ce fait, il est possible de développer une résistance absolue à une attaque parfaitement connue. Par exemple, la compression JPEG étant essentiellement une quantification de coefficients DCT, la prendre en compte dans le choix des quantificateurs permet de



(a) Si $v_i = 0$ (b) Si $v_i = 1$

FIG. 1.6 – Principe du tatouage par quantification scalaire avec des symboles binaires

résister à l'attaque.

L'inconvénient majeur réside dans la résistance aux changements d'échelle du canal lors de l'attaque ($y'_i = \gamma_i \times y_i$) et à l'ajout de bruit non blanc. Les quantificateurs sont décalés et l'extraction est mise en défaut. Néanmoins, une solution proposée par [EBTG02] consistant à insérer un signal pilote permet d'estimer le facteur γ_i. Ce type de synchronisation est inspirée des techniques utilisées pour estimer les attaques géométriques.

Un autre point délicat est qu'il faut que le décodeur connaisse les états de quantification utilisés lors de l'insertion (pour les cas les plus simples, cela se résume au pas de quantification). Il faut donc transmettre ce paramètre.

Un cas particulier de tatouage par quantification est étudié par Chen et Wornell [CW00]. Ils proposent non pas de quantifier les données x, mais ces données projetées sur un sous-ensemble défini par une base orthogonale. Ce schéma, nommé ST-QIM pour *spread transform quantization index modulation*, permet d'obtenir des échantillons dont les variations seront limitées. Une autre amélioration est d'utiliser une *dither quantization* [CW01, EG00], expliquée en détail dans [GN98]. Cela consiste à insérer avant quantification un signal permettant de réduire la distorsion entre les données quantifiées puis reconstruites et les données originales. Il peut de plus servir de clef nécessaire à l'extraction.

Imposer une contrainte à y

De façon plus générale, le tatouage substitutif est vu comme un tatouage par contrainte : en modifiant x, on force les données marquées à respecter certaines propriétés qui déterminent le message transmis. Le tatouage par quantification vu au-dessus se définit par la contrainte : la donnée est modifiée afin de correspondre à un état de quantification codant une partie du message.

Le schéma proposé par Koch et Zhao [KZ95] est une illustration de ce principe. Un triplet de données hôtes (x_1, x_2, x_3) est associé à un bit. Le tatouage consiste à forcer le triplet marqué à respecter une relation d'ordre dépendante du bit à transmettre :

$$\begin{array}{llll} y_1 > y_3 + \delta & y_2 > y_3 + \delta & \text{pour transmettre 1} & \\ y_1 + \delta < y_3 & y_2 + \delta < y_3 & \text{pour transmettre 0,} & (1.10) \end{array}$$

où δ conditionne la robustesse du schéma. L'extraction se fait en vérifiant la relation d'ordre des triplets reçus : si $y'_1 > y'_3$ et $y'_2 > y'_3$, alors le bits extrait est 1. Dans le cas contraire, le bit 0 est extrait. D'autres exemples de tatouage par contrainte sont les techniques inspirées de la compression fractale [PJ96, BCD98] ou celles utilisant une modification des statistiques de sous-bandes issues de la transformée en ondelettes du document hôte [Man01].

Une propriété intéressante du tatouage substitutif est que les données hôtes n'interfèrent pas dans le décodage. En l'absence d'attaque, on est certain de décoder correctement le message inséré, contrairement au tatouage additif.

1.3 Extraction ou détection ?

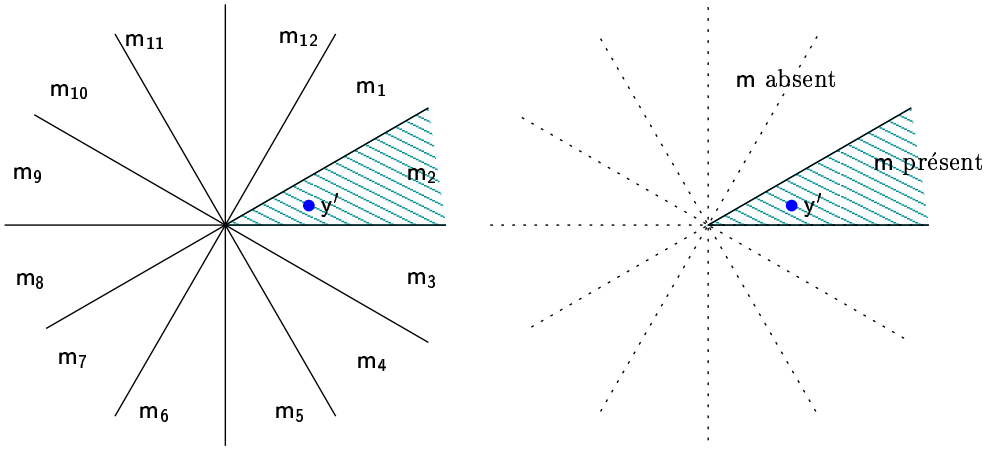
Les techniques présentées ici permettent de transmettre un certain nombre de symboles. À la réception des données y', une suite de symboles est extraite. On transmet donc un message. Une autre application du tatouage est la détection. Dans ce cas, la phase d'extraction indique si la marque est présente au pas.

La figure 1.7(a) illustre l'extraction d'un message. Nous avons vu qu'un message m est associé à une marque w grâce à la fonction de codage. Si l'on suppose que toutes les marques possibles sont de même énergie, les w sont sur une hyper-sphère centrée en $\mathsf{0}$, et sont régulièrement répartis grâce à la fonction de codage et aux propriétés des espaces de grande dimension. L'extraction se fait en recherchant le $\widehat{\mathsf{w}}$ le plus proche du signal y'. On associe à chaque w une zone d'attraction telle que si y' est dans cette zone, la marque w correspondante est la plus proche. Ces zones sont tracées en noir sur la figure. Cela définit des hyper-cônes dans l'espace n-dimensionnel. Un bon schéma de tatouage en vue de la transmission d'un message consiste donc à

- trouver une fonction de codage maximisant le volume de chaque zone de robustesse pour une taille de message donnée (bonne répartition des mots de code),
- en considérant la zone de robustesse associée au message à transmettre, construire un signal marqué tel que les attaques ne le fassent pas quitter cette zone,
- avoir une fonction de décodage permettant de trouver la marque la plus proche parmi les $\|\mathcal{M}^k\|$ possibles en un temps réaliste.

On remarque que même si le signal reçu n'a pas été tatoué, un message sera tout de même extrait.

La figure 1.7(b) illustre la détection d'une marque depuis le signal y'. Contrairement à l'extraction, il y a un seul message possible. On retrouve également une zone de robustesse (ou zone de détection dans ce cas, en vert sur la figure), mais elle est définie par le détecteur. C'est un compromis entre robustesse de la détection et probabilité de fausse alarme : si la zone est étroite, il y a peu de chances qu'un signal non marqué s'y trouve, mais une attaque faible pourra déplacer le signal marqué hors de la zone (et fausser la détection). Un schéma de tatouage pour la détection consiste à :



(a) Extraction d'un message : on obtient $\widehat{\mathsf{m}} = \mathsf{m}_2$

(b) Détection du message m

FIG. 1.7 – Illustration de la différence entre extraction et détection ($n = 2$, $\|\mathcal{M}^k\| = 12$)

- trouver le meilleur compromis entre robustesse et probabilité de fausse alarme,
- construire un signal marqué tel que les attaques ne le fassent pas quitter cette zone, en considérant la zone de robustesse/détection visée
- avoir une fonction permettant de savoir si le signal marqué appartient ou non à la zone de détection en un temps réaliste.

La phase d'insertion de la marque (construction du signal y) est la même pour la transmission ou pour la détection d'un message. Seules les fonctions de codage/décodage changent. Un algorithme de détection peut donc être vu comme un cas particulier de transmission. Les techniques proposées par la communauté scientifique sont utilisables dans les deux cas.

Conclusion

Le tatouage robuste permet de transmettre un message *via* un document hôte. Cette transmission se compose de trois phases : insertion, diffusion et extraction. L'insertion d'une marque modifie le document hôte et introduit donc une distorsion. Le document marqué est ensuite sujet à des modifications lors de sa diffusion. Elles peuvent être involontaires (traitements usuels sur les documents multimedia) ou conçues dans le but de supprimer la marque. Ces modifications, vues comme des attaques sur le système de tatouage, introduisent également une distorsion. Enfin, la phase d'extraction utilise le document marqué et attaqué afin d'extraire le message s'y trouvant. Un système de tatouage est robuste s'il peut résister à de fortes attaques sans que cela mette en péril la bonne transmission du message.

Les méthodes de tatouage peuvent se classer dans deux catégories. Les techniques

additives ajoutent un signal de marque correspondant au message à transmettre, au document hôte. L'extraction consiste alors à retrouver ce signal ajouté afin d'en déduire le message. Les techniques substitutives modifient les données hôtes afin que leur forme ou leurs propriétés correspondent au message souhaité. Lors de la réception du document marqué, l'analyse des données permet de déduire le message. Les techniques additives font que le signal hôte est un bruit limitant la performance de la transmission : même en l'absence d'attaque, il est possible d'avoir des erreurs d'extraction. Les techniques substitutives n'ont pas ce problème.

Chapitre 2

Représentation et modélisation du signal hôte

Comme nous l'avons vu dans le chapitre précédent, les opérations propres au tatouage (insertion, codage et décodage) s'appuient sur une représentation abstraite des données (vecteurs x, y, ...), sans lien avec le type de document. Cela permet de généraliser les méthodes de tatouage à tout type de documents multimédia (images monochromes et en couleurs, vidéos et sons) et à tout type de représentation.

Le moyen le plus simple d'obtenir un vecteur de données x est de considérer directement les valeurs des échantillons du document : il s'agit de tatouage dans le domaine spatial. On peut décomposer le problème en utilisant plusieurs vecteurs de données hôtes si on a affaire à des document multi-composantes, comme par exemple $\mathsf{x} = \{\mathsf{x}_\mathsf{r}, \mathsf{x}_\mathsf{g}, \mathsf{x}_\mathsf{b}\}$ pour des images en couleurs RGB. Or, la représentation spatiale se prête mal à l'analyse perceptuelle et à la modélisation des attaques : il est difficile de prévoir l'impact des attaques sur les données marquées. Ainsi, les compressions telles que celles des normes JPEG [Wal91] ou MPEG Layer 3 (plus connu sous le nom de MP3) travaillent sur une représentation fréquentielle et modifient principalement les hautes fréquences, peu influentes perceptuellement. Dans le cas d'un tatouage dans le domaine spatial, il est difficile d'isoler ces hautes fréquences, qui seront potentiellement plus attaquées et qui nécessitent donc un traitement particulier lors des phases d'insertion et d'extraction. Les transformées fréquentielles permettent d'analyser plus finement le document à marquer et d'obtenir une représentation plus adaptée au tatouage robuste. Nous verrons dans la suite de ce chapitre plusieurs transformées, en nous focalisant sur leurs propriétés statistiques. Nous étudierons ensuite les modèles perceptuels, indispensables pour s'assurer de l'invisibilité ou de l'inaudibilité du tatouage une fois revenu dans le domaine spatial ou temporel.

2.1 Transformées fréquentielles

Les transformées fréquentielles trouvent de nombreuses applications dans la restoration de qualité ou l'analyse [GW92]. Elles sont particulièrement utilisées pour la

compression : la transformée en cosinus discrète est l'élément central des normes de compression JPEG et MPEG, et la transformée en ondelettes est utilisée dans la norme JPEG 2000 [TM01].

2.1.1 Transformée de Fourier

Considérons un signal discret s composé de n échantillons. Sa transformée de Fourier discrète (DFT), notée f est définie par la paire

$$\forall u \in \{1, 2, \ldots, n\},\ f_u = \sqrt{\frac{1}{n}} \sum_{x=1}^{n} s_x \exp\left[-2i\pi \frac{(u-1)(x-1)}{n}\right], \tag{2.1}$$

$$\forall x \in \{1, 2, \ldots, n\},\ s_x = \sqrt{\frac{1}{n}} \sum_{u=1}^{n} f_u \exp\left[2i\pi \frac{(u-1)(x-1)}{n}\right] \tag{2.2}$$

avec $i = \sqrt{-1}$. Ce signal complexe peut s'écrire $f_u = |f_u| \exp[i \times \phi_u]$. La fonction $|\mathsf{f}|$ est le spectre de Fourier de s tandis que ϕ est sa phase angulaire. Le signal $|\mathsf{f}|^2$ est le spectre de puissance de s. Un exemple est donné sur la figure 2.1 pour l'image *Lena*. Il existe un algorithme rapide pour calculer la transformée : la FFT (pour *fast Fourier transform*). Il permet de passer d'une complexité en $\mathcal{O}(n^2)$, comme le suggère l'équation 2.1, à une complexité en $\mathcal{O}(n \log_2 n)$ en calculant f sous forme récursive.

À partir de l'équation 2.1, on peut remarquer qu'une translation sur s ne modifie pas le spectre $|\mathsf{f}|$. Cette invariance du spectre est une propriété intéressante en vue de la résistance à des attaques désynchronisantes. Néanmoins, la rotation ou le changement d'échelle dans le domaine spatial ont le même impact sur le spectre. La transformée de Fourier-Mellin [RP98] s'appuie sur cette caractéristique pour développer une invariance plus complète. Si l'on considére un signal bi-dimensionnel, les coordonnées d'un élément (x, y) peuvent se mettre sous la forme log-polaire $(e^{\mu} \cos\theta, e^{\mu} \sin\theta)$, avec $\mu \in \mathbb{R}$ et $\theta \in [0; 2\pi[$, et on définit une équivalence $(x, y) \leftrightarrow (\mu, \theta)$. Ces coordonnées polaires disposent de deux propriétés importantes. D'abord, le changement d'échelle correspond à une translation :

$$(\Delta \times x, \Delta \times y) \leftrightarrow (\mu + \log \Delta, \theta)\,. \tag{2.3}$$

De la même façon, une rotation d'un angle Δ est également convertie en translation :

$$(x \cos\Delta - y \sin\Delta, x \sin\Delta + y \cos\Delta) \leftrightarrow (\mu, \theta + \Delta)\,. \tag{2.4}$$

En injectant ce système de coordonnées dans la transformée de Fourier (spectre de Fourier passé en coordonnées polaires, suivi d'une autre transformée de Fourier), on définit la transformée de Fourier-Mellin. Le spectre de Fourier-Mellin d'un signal bi-dimensionnel est invariant à la rotation, au changement d'échelle et bien sûr à la translation. Cette transformation définit donc une représentation invariante à ces trois modifications[1], utile pour la résistance aux attaques géométriques. Mais le passage des coordonnées euclidiennes aux coordonnées polaires implique des approximations numériques

[1] Invariance RST.

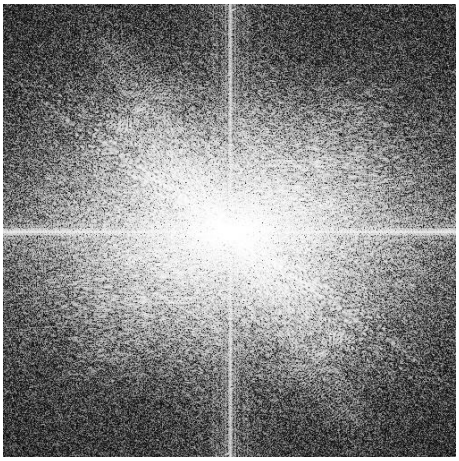


FIG. 2.1 – Spectre de Fourier de l'image *Lena*

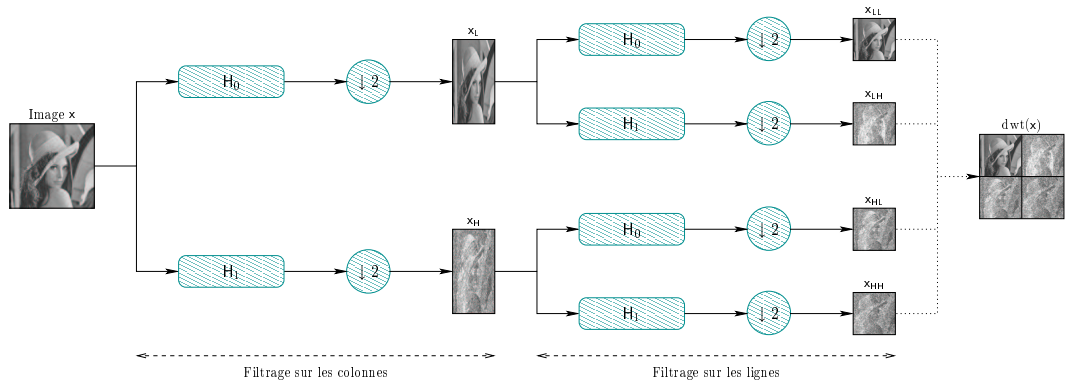
dans le cas discret, liées à un problème de changement de grille d'échantillonnage. Ces approximations font que l'enchaînement transformée de Fourier-Mellin et transformée inverse n'est pas sans perte : la transformée n'est pas parfaitement inversible en pratique et introduit une distorsion supplémentaire si elle est utilisée dans un schéma de tatouage.

2.1.2 Transformée en cosinus

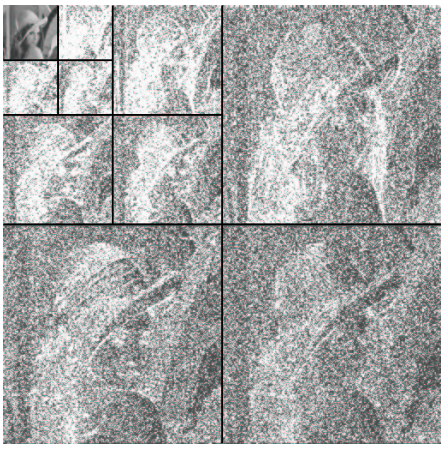
Depuis quelques années, la transformée en cosinus discrète (DCT) est la représentation de choix pour la compression (MP3, JPEG et MPEG), grâce à son compromis intéressant entre pouvoir de décorrélation, proche de l'optimal, et complexité algorithmique. Elle est utilisée dans les algorithmes de compression JPEG et MPEG par blocs de taille 8×8. Cette caractéristique est très souvent reprise dans les techniques de tatouage utilisant la DCT.

2.1.3 Transformée en ondelettes

La transformée en ondelettes discrète (DWT) permet de transformer un signal discret en sous-bandes directionnelles, assimilables à une décomposition fréquentielle. Cette décomposition est récursive, comme illustré par la figure 2.2(b). Outre l'interprétation fréquentielle des sous-bandes, on remarque que la structure spatiale du signal est conservé (au contraire des transformées de Fourier et DCT) : on a un aspect spatio-fréquentiel. De plus, la DWT permet une approche multi-résolution du signal. Cela peut être exploité dans le cadre du tatouage vis-à-vis des changements d'échelle. Cette transformée est la pierre angulaire du récent format de compression d'image JPEG 2000 [TM01].



(a) Principe de la transformée en ondelettes vue par l'approche bancs de filtres, appliqué sur une image



(b) Transformée en ondelettes de l'image *Lena* sur trois niveaux

FIG. 2.2 – Transformée en ondelettes

2.2 Modélisation statistique

La connaissance de la distribution statistique du signal hôte est une donnée indispensable pour modéliser de façon rigoureuse le tatouage. Ainsi, on exprime analytiquement l'impact des attaques sur l'extraction du messsage, les mesures de distorsion et les probabilités d'erreur.

La modélisation la plus générale possible, et au final la plus proche de la réalité, est de considérer un modèle statistique indépendant pour chacun des échantillons considérés. Mais ce modèle est d'une complexité importante et ne peut être utilisé tel quel par de nombreux outils de codage canal utilisables pour le tatouage (développés dans le chapitre suivant). À l'opposé, prendre un modèle unique (gaussienne généralisée pour les coefficients DCT [Mül93, BBRP99], distribution de Weibull pour la transformée de Fourrier [RBB+99], ...) est en contradiction avec l'objectif de décorrélation des transformées fréquentielles.

La solution intermédiaire consiste à regrouper les échantillons de distributions proches, et de considérer chacun de ces sous-groupes comme indépendants et identiquement distribués (i.i.d.). Ainsi, les travaux de Moulin [hKRM99, Mou01] utilisent un modèle gaussien, et par regroupement, définit un ensemble de canaux parallèles gaussiens. Dans le cas de la DWT, la proposition de Baraniuk *et al.* [CNB98] trouve un bon compromis entre simplicité et précision. Pour chaque sous-bande, le modèle considère deux familles de coefficients. Ces familles sont formées en appliquant des règles (les coefficients voisins tendent à être de même famille, une famille se propage à travers les sous-bandes, ...). L'information d'appartenance à une ou l'autre famille est modélisée par un réseau de Markov caché. Un algorithme de type EM (*expectation maximization*) permet de déterminer la probabilité avec laquelle un coefficient appartient à une famille. Chaque sous-bande est alors modélisée par une mixture de deux lois Normales : $\Pr(x_i \sim \mathcal{N}_1) \times \mathcal{N}_1 + \Pr(x_i \sim \mathcal{N}_2) \times \mathcal{N}_2$. Une approche assez similaire est donnée par Weidmann et Vetterli [WV99]. Ils répartissent les coefficients DWT en deux catégories : les significatifs (énergie forte) et les non-significatifs (énergie faible). Chaque catégorie est modélisée par une loi Normale.

2.3 Mesures et modèles perceptuels

La définition même du tatouage indique que les modifications apportées au document hôte doivent rester imperceptibles, ou tout au plus ne pas gêner l'exploitation normale du document marqué. Afin de respecter cette condition, ou de pouvoir mesurer de façon efficace la distorsion introduite par le tatouage, il est nécessaire d'introduire un critère perceptuel basé sur une modélisation de la perception des signaux multimedia.

La mesure habituellement utilisée pour quantifier la distorsion entre un signal original `x` et un signal modifié `y` est le PSNR[2]. Elle est basée sur l'erreur quadratique

[2]Pour *peak signal-to-noise ratio*.

moyenne[3], définie par

$$\mathrm{eqm}(\mathsf{x},\mathsf{y}) = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - y_i\right)^2, \tag{2.5}$$

où n est la dimension commune aux deux vecteurs considérés. Quant au PSNR, il est calculé par

$$\mathrm{psnr}(\mathsf{x},\mathsf{y}) = 10\log_{10}\left[\frac{\left(\max(\mathsf{x})\right)^2}{\mathrm{eqm}\,(\mathsf{x},\mathsf{y})}\right]. \tag{2.6}$$

On peut remarquer que chaque échantillon de x intervient de la même façon dans le calcul. Or, cela n'est pas le cas lorsque l'on considère le système perceptuel humain. La figure 2.3 illustre le fait que pour une même mesure de PSNR, les résultats peuvent être visuellement de qualités bien différentes.

À l'opposé, le critère le plus fidèle à la perception est la mesure de qualité subjective, comme le MOS (*mean opinion score*) : les documents sont notés par un panel de testeurs et la moyenne de ces notes donne une mesure de qualité. Mais il est évident que ce type de mesure est extrêmement contraignante en temps et en moyens. Dans le cas du tatouage, la mesure de qualité doit être mesurée au moment même de la phase d'insertion. La suite de cette section va donc s'intéresser aux métriques objectives.

2.3.1 Exemple de caractéristiques perceptuelles : le système visuel humain

La sensibilité du système visuel humain (HVS) dépend principalement de trois paramètres : la fréquence spatiale, la couleur et l'intensité lumineuse (la luminosité). La réponse perceptuelle en fonction de la fréquence spatiale correspond à la sensibilité au contraste. Cette réponse définit une fonction notée CSF[4], étudiée dans [MS74]. Le HVS est sensible aux contrastes moyens, et peu stimulé par les contrastes très forts ou très faibles. De plus, la sensibilité varie selon l'orientation de cette fréquence : l'œil est plus sensible aux motifs horizontaux et verticaux, plutôt qu'aux motifs à 45 degrés. Le second paramètre est la fréquence spectrale, c'est-à-dire la couleur. Le HVS n'est en effet pas sensible de la même manière aux différentes longueurs d'onde du spectre visible. Dans le cas d'une représentation de la couleur sous la forme de trois canaux {rouge, vert, bleu}, le canal bleu est celui qui a le moins d'importance (le HVS y est moins sensible). Enfin, le dernier paramètre est la luminosité. L'œil peut remarquer de plus petites variations de luminosité quand la luminosité moyenne est faible.

À ces trois paramètres s'ajoute un phénomène plus spécifique : les effets de masquage. La plus ou moins bonne perception d'une modification dépend de son contexte. Ainsi, un élément parfaitement perçu isolément peut être difficilement discernable parmi d'autres éléments de même fréquence, mais de luminosité plus importante. Le masquage peut être fréquentiel (une fréquence en masque une autre) ou fonction de la luminosité. Ces caractéristiques doivent être prises en compte dans les modèles perceptuels

[3]EQM, ou MSE pour *mean square error*.
[4]Pour *constrast sensitivity function*.

(a) Image originale x : *Lena*

(b) Image y_1 : augmentation de constraste. Extrait de [WBL02]

(c) Image y_2 : compression au format JPEG. Extrait de [WBL02]

FIG. 2.3 – Différents résultats visuels pour une même mesure de PSNR : $\mathrm{psnr}(\mathsf{x}, \mathsf{y}_1) \simeq \mathrm{psnr}(\mathsf{x}, \mathsf{y}_2) \simeq 25$ dB

afin qu'ils soient proches de la réalité. Les démonstrations conçues par Adelson [Ade] à partir de ses travaux [Ade00] illustrent ces caractéristiques par des illusions d'optique.

2.3.2 Pondération perceptuelle

L'approche la plus pratique pour prendre en compte les caractéristiques vues précédemment est l'introduction d'une pondération perceptuelle au sein de la mesure classique du PSNR ou de l'erreur quadratique moyenne. Ce type de mesure, noté *w*PSNR par [VDP00] pour *weighted* PSNR, est défini par

$$\mathrm{wpsnr}(\mathsf{x},\mathsf{y}) = 10\log_{10}\left[\frac{(\max(\mathsf{x}))^2}{\mathrm{weqm}(\mathsf{x},\mathsf{y})}\right] \tag{2.7}$$

$$\mathrm{weqm}(\mathsf{x},\mathsf{y}) = \frac{1}{n}\sum_{i=1}^{n}\varphi_i^2\,(x_i-y_i)^2 \tag{2.8}$$

où φ_i est une pondération représentant l'importance du $i^{\text{ème}}$ échantillon. De nombreuses pondérations de ce type ont été proposées, telles que celles citées par [WBL02]. Néanmoins, la plus connue est celle de Watson [Wat93], définie pour les images passées dans le domaine DCT[5], dont une version simplifiée est utilisée dans JPEG 2000 [Tau00] :

$$\varphi_i^2 \propto \frac{1}{\sigma_{b_i}^2+V_i^2} \tag{2.9}$$

$$\text{avec } V_i = \frac{1}{\|\Phi_i\|}\sum_{j\in\Phi_i}|x_j|^{\rho}. \tag{2.10}$$

L'ensemble Φ_i de taille $\|\Phi_i\|$ représente les indices des voisins du $i^{\text{ème}}$ coefficient. Les meilleurs résultats sont obtenus pour $\rho = 1/2$. La valeur V_i est une mesure d'activité au voisinage du $i^{\text{ème}}$ coefficient. La variable σ_{b_i} est un seuil de visibilité, dépendant de la distance d'observation et fixé à 10^{-2} pour JPEG 2000. La version complète de la mesure de Watson permet de prendre en compte la sensibilité fréquencielle (une table donne le niveau de sensibilité pour les 64 coefficients d'un bloc DCT) et les phénomènes de masquage dûs à la luminance et au contraste.

Les travaux de l'équipe de T. Pun [VHBP99] s'appuient sur une pondération calculée à partir d'une constatation simple : l'œil agit comme un filtre débruiteur. Plus le filtre supprime de bruit, et moins le HVS sera sensible à ce bruit. Ainsi, dans le cas de signaux non i.i.d. et gaussiens (chaque échantillon x_i est modélisé par $X_i \sim \mathcal{N}(0,\sigma_{X_i}^2)$), le filtre optimal au sens du MAP est le filtre de Wiener. Les auteurs définissent alors une mesure, notée NVF pour *noise visibility function*, d'une forme similaire à celle de la pondération d'un filtre de Wiener :

$$\mathrm{nvf}(i) = \frac{1}{1+\sigma_{X_i}^2}. \tag{2.11}$$

[5] À l'origine, Watson voulait utiliser sa pondération au sein de JPEG, qui se base sur une quantification de blocs DCT de taille 8×8.

La variance locale est calculée en pratique en utilisant une fenêtre 3×3. L'autre cas considéré prend en compte une modélisation par gaussienne généralisée de moyenne $\overline{x}$, de variance σ_X^2 et de facteur de forme γ. En calculant le filtre débruiteur optimal, le NVF est défini par

$$\begin{aligned} \mathrm{nvf}(i) &= \frac{w_i}{w_i + \sigma_{X_i}^2} \\ \text{avec } w_i &= \frac{\gamma}{||r_i||^{2-\gamma}} \\ \text{et } r_i &= \frac{x_i - \overline{x}}{\sigma_X}. \end{aligned} \tag{2.12}$$

Il peut être utilisé pour définir une mesure de type wPSNR en prenant $\varphi_i = \mathrm{nvf}(i)$.

2.3.3 Seuils de perception

Une autre façon de prendre en compte le système perceptuel humain est d'utiliser des seuils de perception. Contrairement aux critères de qualité sous forme de pondérations vus précédemment, ce type de seuil ne mesure pas une distortion, mais indique la distorsion maximale autorisée sans que la modification soit visible ou audible. Au dessous de ce seuil, la modification ne pourra pas être remarquée, mais au dessus elle pourra être perçue. Ce niveau de distorsion maximal est noté JND (pour *just noticeable difference*). Watson [WYSV97] a déterminé expérimentalement des seuils de perception du bruit pour les coefficients DWT, utilisés afin de calculer des matrices de quantification pour faire de la compression d'images.

Ce type de mesure ne permet pas de quantifier la distorsion perceptuelle introduite : on sait juste si la modification apportée est perceptible ou non. Or, rester sous ce seuil de visibilité impose des modifications très faibles. Appliqué au tatouage, le respect de la JND ne permet donc pas d'introduire une marque de forte énergie. Cela peut poser problème dans le cas où la robustesse est primordiale, et plus importante que la perte de qualité (on peut supporter dans certains cas une dégradation légèrement visible). De plus, ne pas dépasser la JND impose pratiquement l'énergie du tatouage : comme il est impossible de quantifier la distorsion (qui sera imperceptible si l'on reste sous le seuil) et comme le tatouage se doit d'être robuste (et donc son énergie doit être maximisée), la meilleure stratégie est d'imposer l'énergie de la marque égale à la JND [Del00]. L'utilisation d'une fonction de visibilité rend donc impossible le fait d'avoir des stratégies de marquage conciliant plusieurs compromis entre distorsion perceptuelle et énergie insérée.

Conclusion

Afin d'être indépendant du type de document hôte (images, sons, ...), nous l'abstrayons en un signal noté `x`. Celui-ci est obtenu directement à partir des données du documents (valeurs des pixels ou des échantillons), mais la plupart des techniques de

tatouage passent par une transformée. Les plus populaires sont la transformée de Fourier, la DCT ou la transformée en ondelettes, souvent utilisées dans l'analyse de signaux (notamment dans la compression avec perte).

Les méthodes de tatouage doivent s'adapter aux données hôtes. Comme il est impossible de construire une technique de tatouage pour chaque occurrence de signal hôte possible, il est indispensable de s'appuyer sur une modélisation statistique de ces signaux. De plus, afin de garantir l'imperceptibilité du tatouage, il faut également connaître et mesurer la distorsion introduite par l'insertion de la marque. Les modèles perceptuels permettent de quantifier l'impact visuel ou auditif d'une modification, généralement en s'appuyant sur les statistiques des données hôtes.

Chapitre 3

Analogie avec le codage canal

Comme vu dans la première section de cette partie, le tatouage consiste à transmettre un message. Cette transmission est bruitée par le signal hôte dans le cas du tatouage additif et par les éventuelles attaques. Nous sommes donc en face d'un problème de communication et de codage canal. Ce rapprochement permet de reprendre les outils développés dans le cadre du codage canal et de modéliser le problème du tatouage, et par la suite de donner des bases et outils théoriques permettant de définir et d'approcher les performances optimales. Nous verrons dans ce chapitre différents types de canaux classiques et les résultats s'y rapportant. Puis nous rappellerons les principes de codage permettant des performances optimales. Enfin nous étudierons un modèle de canal particulièrement bien adapté au tatouage : les canaux avec information adjacente disponible à l'encodage.

3.1 Canaux pour le tatouage

3.1.1 Canal général

Un **canal discret** est un système consistant en un alphabet d'entrée $\mathcal{W}$, un alphabet de sortie $\mathcal{Y}$ et une matrice de probabilités de transition $\Pr(Y = y \,|\, w)$ qui exprime la probabilité d'observer le symbole y en sortie alors que le symbole w à été envoyé. Un canal est **sans mémoire** si la distribution de probabilités de la sortie dépend uniquement de l'entrée au même moment, et est conditionnellement indépendante des entrées et sorties précédentes [CT91]. La **capacité** d'un canal est le plus grand taux de bits d'information que l'on peut envoyer par utilisation du canal, avec une probabilité d'erreur aussi faible que l'on veut. Elle est donnée par

$$C = \max_{\Pr(w)} \{I(W;Y)\}, \tag{3.1}$$

où $I(W;Y)$ est l'information mutuelle entre W et Y, variables aléatoires modélisant respectivement les signaux w et y, et où le maximum est pris parmi toutes des distributions possibles $\Pr(w)$.

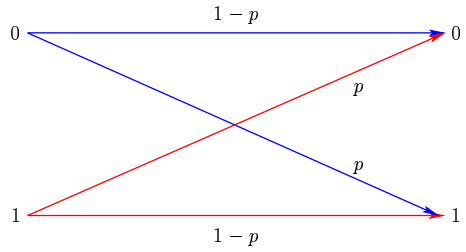


FIG. 3.1 – Transmission d'un symbole sur un canal binaire symétrique

Un exemple simple est le canal binaire symétrique, où $\mathcal{W} = \{0, 1\}$. À chaque utilisation du canal, l'émetteur transmet le symbole 0 ou 1. Le récepteur reçoit le symbole transmis avec une probabilité $1 - p = 1 - \Pr(Y \neq w)$ (figure 3.1). On peut calculer la capacité :

$$\begin{aligned} C &= \max_{\Pr(w)} \{I(W;Y)\} \\ &= \max_{\Pr(w)} \{H(Y) - H(Y \,|\, W)\} \\ &= \max_{\Pr(w)} \{H(Y)\} - H(p) \\ &= 1 - H(p), \end{aligned} \tag{3.2}$$

avec $H(Y)$ entropie de Y. Le maximum $H(Y) = 1$ est obtenu pour une distribution uniforme de W.

3.1.2 Canal gaussien

Considérons un signal à transmettre w borné en énergie par P : $\sum_{i=1}^{n} w_i^2 \leq nP$. Ce signal est bruité par z, bruit modélisé par une loi Normale de moyenne nulle et d'énergie N. Le signal reçu est donc $\mathsf{y} = \mathsf{w} + \mathsf{z}$. Cela est résumé par la figure 3.2. Cette transmission définit un canal gaussien caractérisé par son rapport signal-à-bruit $E_b/N_0 = P/N$.

Dans le cas de symboles binaires, la meilleure stratégie consiste à transmettre $w_i = \pm\sqrt{P}$ [CT91]. Le signal reçu, dont chaque symbole y_i est modélisé par la variable aléatoire Y_i, aura donc la forme de deux gaussiennes centrées en $-\sqrt{P}$ et $+\sqrt{P}$ et de variance N. La loi de décodage optimale correspondant à ce type de transmission, en considérant chaque symbole reçu séparément, est de décoder $\widehat{w}_i = +\sqrt{P}$ si $y_i > 0$ et $\widehat{w}_i = -\sqrt{P}$ sinon. La probabilité d'erreur par bit est alors

$$\mathbf{P}_{\mathrm{e}}^{\mathrm{b}} = \frac{1}{2}\left[\Pr\left(Y_i < 0 \,|\, w_i = +\sqrt{P}\right) + \Pr\left(Y_i > 0 \,|\, w_i = -\sqrt{P}\right)\right] \tag{3.3}$$

$$= \frac{1}{2} \times \operatorname{erfc}\left[\sqrt{\frac{E_b}{2N_0}}\right]. \tag{3.4}$$

La capacité théorique maximale est obtenue en prenant un signal envoyé suivant une loi Normale ($W \sim \mathcal{N}(0, P)$), et est donnée par

$$\begin{aligned} C &= I(W;Y) && (3.5)\\ &= \frac{1}{2}\log_2\left[1+\frac{E_b}{N_0}\right], && (3.6)\end{aligned}$$

où $I(W;Y)$ est l'information mutuelle entre W et Y, variables aléatoires modélisant respectivement les signaux w et y.

Ce type de canal gaussien, aussi nommé AWGN [1], est souvent utilisé pour représenter le tatouage [SPR98, MO99, MO00, CL02] : l'attaque est modélisée comme l'ajout de bruit gaussien. Ce cas est valable uniquement pour les attaques non-désynchronisantes. Néanmoins, même dans ces conditions, de nombreux traitements ne peuvent être considérés comme un simple ajout de bruit. Par exemple, un filtrage passe-bas ou une compression de type JPEG suppriment ou atténuent certaines composantes fréquentielles, agissant comme un bruit multiplicatif corrélé.

3.1.3 Canal SAWGN

Un modèle plus général, considérant à la fois le bruit multiplicatif et additif est considéré dans les articles les plus récents [SEG01a, SEG01b, MI03] : c'est le canal SAWGN[2]. Un signal w transmis *via* un canal SAWGN est reçu comme

$$\mathsf{y} = \gamma \times \mathsf{w} + \mathsf{z}, \tag{3.7}$$

où γ est un facteur multiplicatif et z un bruit additif suivant une loi gaussienne $\mathcal{N}(0, N)$, comme vue au-dessus. En se ramenant à un canal AWGN équivalent, il est facile de montrer que la capacité d'un tel canal est donnée par

$$C = \frac{1}{2}\log_2\left[1+\frac{\gamma E_b}{N_0}\right]. \tag{3.8}$$

Ce type de canal permet d'enrichir la gamme des attaques non-désynchronisantes que peut subir un signal marqué, mais il a également été montré [SG99, SEG01b] que pour une même distorsion, les attaques de la forme de l'équation 3.7 sont plus néfastes pour le tatouage que les attaques AWGN.

3.2 Codage du message

Les capacités théoriques vues dans la section précédente ne sont atteignables qu'avec un codage adapté du message à transmettre. Transmettre le message bit à bit, même si la capacité du canal est supérieure à la taille du message, expose chacun des bits à

[1] Pour *additive white Gaussian noise*.
[2] Pour *scaling and additive white Gaussian noise*.

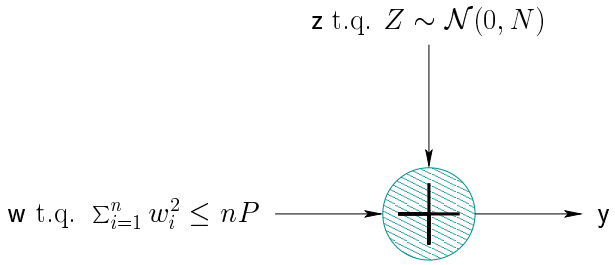


FIG. 3.2 – Transmission d'un signal w *via* un canal gaussien

une probabilité d'erreur non négligeable (équation 3.4), rendant la réception correcte de l'ensemble du message peu probable : $\mathbf{P}_{\mathrm{e}} = 1 - \left(1 - \mathbf{P}_{\mathrm{e}}^{\mathrm{b}}\right)^n$. Un code correcteur permet de limiter ces erreurs. Nous verrons d'abord leur principe, puis leur utilisation pour le tatouage.

3.2.1 Principes

La correction d'erreur se fait en ajoutant de la redondance au message à coder. Dans le cas binaire, à partir d'un message de k bits, le codeur génère un mot de code de n bits. C'est donc une fonction

$$\begin{array}{rcl} \{0,1\}^k & \longrightarrow & \mathcal{U} \subset \{0,1\}^n \\ f : \mathsf{m} & \longmapsto & \mathsf{u}. \end{array} \tag{3.9}$$

Le rendement du code est défini par $r = k/n$. On mesure sa performance par la distance minimale $d_{\min}$ du code, c'est-à-dire la distance entre les deux éléments de $\mathcal{U}$ les plus proches. La correction d'erreur (le décodage) consiste à retrouver le mot de code $\mathsf{u}^\star \in \mathcal{U}$ le plus proche du signal reçu. La fonction $f^{-1}()$ donne alors le message correspondant. On voit ainsi pourquoi la distance minimale intervient dans le pouvoir de correction du code : si le signal à décoder est séparé du bon mot de code par une distance supérieure à $d_{\min}/2$, il est possible que le mot de code le plus proche ne soit pas celui initialement transmis. D'un point de vue géométrique, dans un espace n-dimensionnel, les mots de codes définissent des hyper-sphères de robustesse de rayon $d_{\min}/2$. Lorsque le signal quitte la sphère, il peut être mal décodé. L'apport d'un code sur la probabilité d'erreur par bit est évaluée [Des01] grâce à

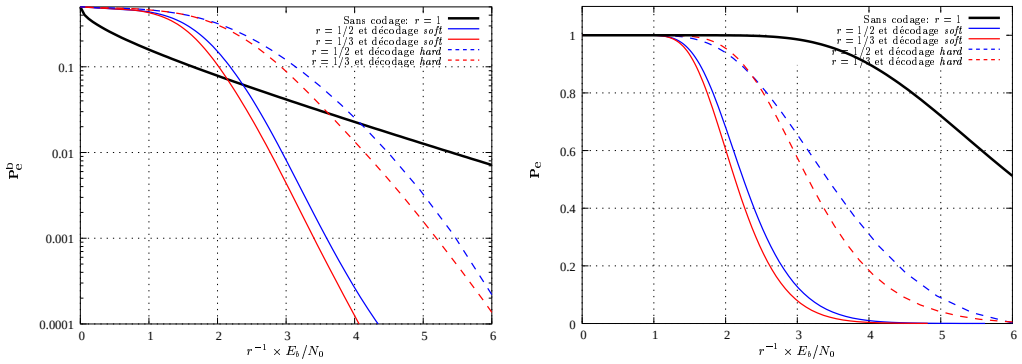
$$\mathbf{P}_{\mathrm{e}}^{\mathrm{b}} \simeq \frac{1}{2} \operatorname{erfc}\left[\sqrt{d_{\min} \frac{E_b}{2N_0}}\right]. \tag{3.10}$$

La recherche de l'élément le plus proche dans un ensemble de grande dimension est un problème algorithmique qui, pour des éléments non ordonnés, n'a pas de meilleure solution que la recherche séquentielle exhaustive. Décoder un message peut donc être très complexe (en $\mathcal{O}(2^k)$). Cela n'est réalisable que pour des messages de petite taille.

De plus, seule une répartition uniforme des mots de codes peut permettre d'atteindre les limites de capacité vues précédemment. Des solutions sous-optimales plus rapides ont donc été développées. La première est de considérer des codes en bloc, tels que les codes BCH ou de Reed-Solomon. Le message est divisé en blocs, codés indépendamment, et le tout est concaténé pour former le mot de code u. Le décodage se fait également par bloc, avec des recherches dans des espaces réduits. La seconde solution permettant d'obtenir une bonne distance minimale et une vitesse de décodage acceptable est donnée par les codes convolutifs. L'encodage se fait grâce à des registres à décalage, alimentés par les bits du message. À chaque top d'horloge, les registres se décalent et r^{-1} bits sont générés par combinaison linéaire des éléments des registres. Cette technique permet d'obtenir un ensemble $\mathcal{U}$ ordonné par la corrélation entre les bits des mots de code. L'algorithme de Viterbi [Vit67] exploite cette propriété par la construction d'un graphe d'états (un treillis convolutif). Chaque chemin de ce graphe correspond à un mot de code possible et la recherche consiste à trouver le chemin minimisant l'erreur avec le signal observé. La complexité est linéaire (en $\mathcal{O}(k)$).

Un bond en avant dans les performances a été fait en 1993 par les turbo-codes [BGT93]. Cette technique, appliquable aux deux types de codes vus au-dessus, utilise deux codeurs différents. Le résultat du décodage de l'un sert d'*a priori* à l'autre, de façon itérative. Les codes LDPC (*low-parity density-check codes*) permettent également d'atteindre d'excellentes performances. Ils ont été initialement inventés dans les années 1960 [Gal63], mais ont eu peu de succès à l'époque à cause de leur complexité importante et de l'arrivée des codes de Reed-Solomon. Ils ont été redécouverts à la fin des années 1990 et depuis, leurs performances et les progrès en terme de capacité de calcul les ont remis à l'ordre du jour.

De façon générale, il y a deux moyens de décoder un message reçu. La première est de convertir l'estimation obtenu en symbole (en choisissant bien sûr le plus proche) et de décoder cette suite de symboles. C'est un décodage avec décision dure (*hard decoding*). L'autre moyen est de directement envoyer au décodeur les estimations : c'est le décodage avec décision souple (*soft decoding*). Ce dernier offre de meilleures performances. Supposons une fonction de codage simple telle que $f(0) = \{+1\ -1\ +1\}$ et $f(1) = \{-1\ +1\ -1\}$. Le symbole 1 est envoyé, c'est-à-dire le mot de code $\{-1\ +1\ -1\}$. À la réception, on estime la suite $\{+0,1\ +0,9\ +0,2\}$. Avec le décodage *hard*, il y a prise de décision sur les symboles recu, et c'est donc la suite $\{+1\ +1\ +1\}$ qui est décodée. Dans ce cas, le mot de code le plus proche est celui correspondant à 0. Il y a donc une erreur de transmission. Avec le décodage souple, le décodeur choisi le mot de code le plus proche de $\{+0,1\ +0,9\ +0,2\}$: c'est celui qui correspond à 1. La transmission est correcte. La figure 3.3 (issue de nos expérimentations) montre la supériorité du décodage souple pour un codeur convolutif associé à un décodeur de Viterbi (courbes en trait plein à comparer à celles de même couleur en pointillés).



(a) Probabilités d'erreur par bit

(b) Probabilités d'erreur par message

FIG. 3.3 – Probabilités d'erreur obtenus avec ou sans codage ($k = 100$, codes convolutifs avec longueur de contrainte égale à 9)

3.2.2 Application au tatouage

Une particularité du canal formé par le tatouage est que l'énergie maximale du signal transmis (la distorsion introduite) est limitée. De ce fait, insérer $n = k/r$ bits au lieu de k bits réduit l'énergie par bit d'un facteur r^{-1}, et donc d'autant le rapport signal-à-bruit E_b/N_0 du canal. On a donc l'estimation de probabilité d'erreur par bit

$$\mathbf{P}_\mathrm{e}^\mathrm{b} \simeq \frac{1}{2}\,\mathrm{erfc}\left[\sqrt{d_\mathrm{min}\frac{r \times E_b}{2N_0}}\right]. \tag{3.11}$$

On peut en déduire que le codage par simple répétition ($d_\mathrm{min} = r^{-1}$) n'apporte pas de gain. Au contraire, si $d_\mathrm{min} > r^{-1}$, un gain de performance est apporté. Cela est visible sur la figure 3.3 où le rapport E_b/N_0 en abscisse est normalisé par le rendement du code, pour souligner l'importance de la distance minimale du code.

Plusieurs études ont mis en évidence l'apport de codes correcteurs dans le tatouage. Ainsi Baudry *et al.* [BDS+01] ont étudié l'intérêt des codes BCH. Pérez-Gonzàles *et al.* [PGHB01] ont proposé une étude poussée sur les codes de Hamming, les codes BCH et les codes convolutifs, dans le cadre d'une décision *hard* ou *soft*. Ils montrent que les codes convolutifs apportent le meilleur compromis entre performance et complexité.

3.3 Canaux avec information adjacente

Les canaux AWGN et SAWGN permettent, en modélisant les attaques sous la forme d'ajout de bruit et de facteur d'échelle, de dériver des formules de capacité. Néanmoins, la caractéristique fondamentale du tatouage est que le message est transmis *via* un signal hôte, intervenant dans les caractéristiques du canal. Dans le cas d'un signal hôte

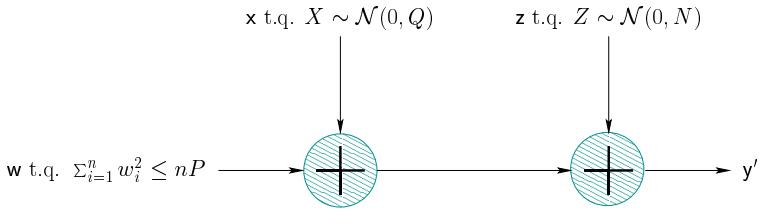


FIG. 3.4 – Transmission d'un signal w *via* un canal gaussien avec information adjacente disponible à l'encodage

x modélisé par $X \sim \mathcal{N}(0, Q)$, le signal reçu est donné par $\mathsf{y}' = \mathsf{x} + \mathsf{w} + \mathsf{z}$ et la capacité est, d'après l'équation 3.6

$$C = \frac{1}{2} \log_2 \left[1 + \frac{P}{Q + N} \right]. \tag{3.12}$$

Or, ce bruit est parfaitement connu au moment de l'encodage permettant d'obtenir w. Ce type de canal a été étudié par Costa [Cos83], qui en a déduit une limite de capacité étonnante : le bruit présent à l'encodage n'influe pas sur la capacité du canal. Il exhibe un schéma de codage théorique permettant d'atteindre cette capacité, présenté dans la suite de cette section.

3.3.1 Schéma de Costa

Les travaux de Gel'fand et Pinskert [GP80] et de Heegard et El Gamal [EH83] montrent que la capacité d'un canal sans mémoire avec l'information x disponible à l'encodeur est définie par

$$C = \max_{\Pr(u,w \,|\, x)} \left\{ I(U;Y') - I(U;X) \right\}. \tag{3.13}$$

avec U variable aléatoire représentant le dictionnaire utilisé. Costa résout cette maximisation en identifiant la forme optimale de ce signal : $\mathsf{u} = \mathsf{w} + \alpha\mathsf{x}$, et donc $U \sim \mathcal{N}(0, P + \alpha^2 Q)$. La capacité du canal est alors définie par

$$C = \max_{\alpha} \left\{ I(U;Y') - I(U;X) \right\} \tag{3.14}$$

$$\text{avec } I(U;Y') = \frac{1}{2} \log_2 \left[\frac{(P + Q + N)(P + \alpha^2 Q)}{PQ\,(1 - \alpha)^2 + N\,(P + \alpha^2 Q)} \right] \tag{3.15}$$

$$\text{et } I(U;X) = \frac{1}{2} \log_2 \left[\frac{P + \alpha^2 Q}{P} \right]. \tag{3.16}$$

Le maximum est obtenu pour $\alpha = P/(P + N)$, donnant

$$C = \frac{1}{2} \log_2 \left[1 + \frac{P}{N} \right]. \tag{3.17}$$

De cette démonstration découle le schéma de Costa (connu sous le nom d'ICS pour *ideal Costa scheme*). Le signal u est un mot de code issu d'un dictionnaire $\mathcal{U}$ composé de $2^{n(I(U;Y')-\epsilon)}$ éléments, avec ϵ tendant vers zéro quand n tend vers l'infini. Le dictionnaire $\mathcal{U}$ est divisé en $2^{n(C-\epsilon)}$ sous-dictionnaires $\mathcal{U}_{\mathsf{m}}$: à chaque message possible m est associé un sous-dictionnaire. Au moment de l'encodage de m, le mot de code $\mathsf{u}^\star \in \mathcal{U}_{\mathsf{m}}$ tel que $\mathsf{u}^\star$ et x soient typiquement joints est recherché. La formule du signal transmis est donnée par

$$\begin{aligned} & \mathsf{w} = \mathsf{u}^\star - \alpha \mathsf{x} \\ \Leftrightarrow \quad & \mathsf{y} = (1-\alpha)\,\mathsf{x} + \mathsf{u}^\star. \end{aligned} \tag{3.18}$$

Le décodage se fait en recherchant le mot de code $\widehat{\mathsf{u}} \in \mathcal{U}$ tel que celui-ci et y' soient typiquement joints. Le sous-dictionnaire $\mathcal{U}_{\widehat{\mathsf{m}}}$ auquel il appartient donne le message décodé $\widehat{\mathsf{m}}$. La formulation de l'équation 3.18 met en évidence le principe essentiel du schéma de Costa. Plutôt que d'ajouter un mot de code potentiellement orthogonal au bruit déjà présent[3], ce bruit est dirigé vers le mot de code $\mathsf{u}^\star/\alpha$ correspondant au message à transmettre, sans toutefois l'atteindre. La taille du dictionnaire et sa subdivision, ainsi que le paramètre α sont calculés afin que statistiquement, le signal w soit d'énergie suffisante pour que y puisse être placé à l'intérieur de la zone de robustesse de $\mathsf{u}^\star$, et qu'il faille y ajouter un bruit d'énergie au moins égale à N pour en sortir.

La difficulté pratique de ce schéma est la construction du dictionnaire structuré, avec des mots de code et des sous-dictionnaires bien répartis. Costa utilise dans sa démonstration un dictionnaire aléatoire afin de s'assurer de cette bonne répartition (grâce à l'*asymptotic equiparition property* [CT91]), mais sa taille rend toute mise en pratique irréaliste.

3.3.2 Propositions pratiques

Malgré le gain potentiel de performance possible grâce au schéma de Costa, une application directe en grande dimension (la capacité théorique étant atteinte quand n tend vers l'infini) est impossible de par sa complexité. Plusieurs approches pratiques, portant principalement sur la construction d'un dictionnaire structuré, plus ou moins sous-optimales, ont été proposées afin de s'approcher de la limite de l'équation 3.17.

Modification du tatouage par quantification

Le tatouage par quantification (section 1.2.2) peut être vu comme un tatouage avec prise en compte de l'information adjacente. En effet, le signal hôte est modifié de façon à correspondre à un état de quantification, c'est-à-dire un mot de code. Cela définit un dictionnaire réparti uniformément dans l'espace (mono-dimensionnel pour la quantification scalaire) structuré en affectant un message à chaque mot de code, comme illustré par la figure 3.6 pour un code à deux dimensions.

[3] Soit une insertion du type $\mathsf{x} + \mathsf{w}$ avec w issu du codage du message sans prise en compte de x.

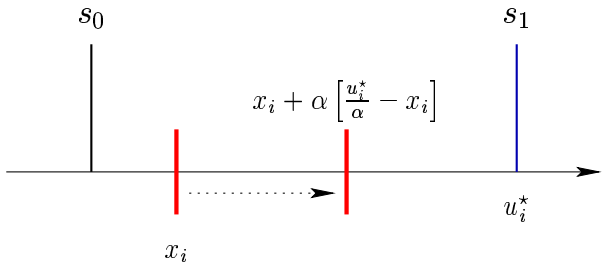


FIG. 3.5 – Principe du SCS, basé sur une quantification scalaire du signal hôte

Dans le tatouage par quantification classique, le signal y tatoué est le mot de code (l'état de quantification) le plus proche des données hôtes. Or, un des principes de l'ICS est de rapprocher le signal hôte vers le mot de code le plus proche, grâce au paramètre α de l'équation 3.18, plutôt que de lui imposer une certaine valeur. Fort de cette constatation, le SCS (*scalar Costa scheme*) de Eggers *et al.* [EBTG02] introduit ce principe dans son précédent schéma de tatouage par quantification [EG00], comme illustré par la figure 3.5. Si cela augmente les performances du tatouage par quantification classique, les réserves émises dans la section 1.2.2 sur ce type de technique sont toujours d'actualité : pas de résistance vis-à-vis du bruit multiplicatif, et besoin de connaître le pas de quantification à l'extraction.

Dictionnaires structurés issus de codes correcteurs

Malgré l'apparente adéquation des codes issus d'une lattice dans le cadre des canaux avec information adjacente, ils ne sont pas de même énergie. Cette lacune les rend sensibles au changement d'échelle du canal. De plus, la bonne répartition des mots de code dans l'espace n'est pas assurée. Une autre solution pour la construction de codes structurés est de s'inspirer des codes correcteurs d'erreurs. En effet, les propriétés recherchées par Costa sont les mêmes qu'en codage canal.

Ainsi, Miller *et al.* [MDC02] utilisent un treillis modifié pour obtenir un dictionnaire, illustré par la figure 3.7. Un treillis est un graphe caractérisé par un nombre d'états 2^r et un nombre de symboles k. Il comporte $k \times 2^r$ sommets. Dans le cas binaire, un treillis classique tel qu'utilisé dans les codes correcteurs d'erreurs comporte deux arcs par sommet (un par bit), comme sur la figure 3.7(a). À une suite de bits (un message) correspond donc un chemin unique. L'idée de Miller est d'avoir plusieurs chemins possibles par message. Pour cela, il multiplie le nombre d'arcs : pour un sommet (s, t) et un bit donné, il y a i arcs possibles vers un sommet de $t + 1$ (figure 3.7(b) où $i = 2$). Et donc le graphe comporte i^k chemins possibles par message, comme le montre la figure 3.7(c).

Ces arcs sont valués par des vecteurs de symboles pseudo-aléatoires. Un chemin correspond donc à une suite de symboles. Lors de l'encodage du message m, le chemin le plus proche des données hôtes x est choisi entre les i^k possibles (en n'autorisant que

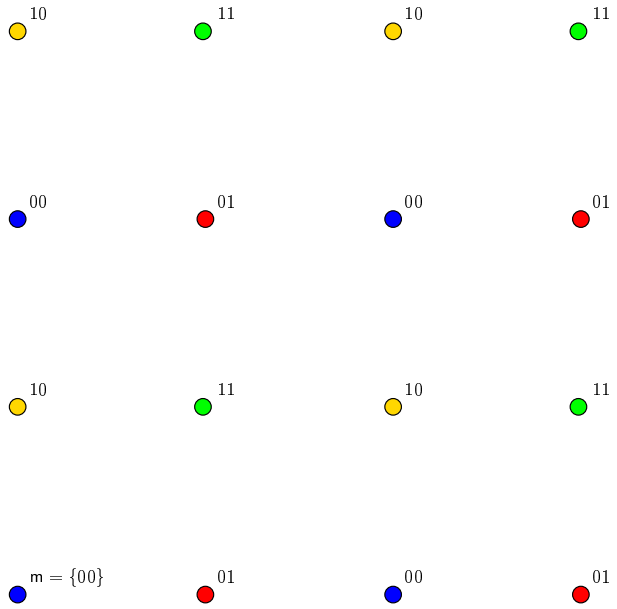


FIG. 3.6 – Construction d'un dictionnaire structuré en utilisant des quantificateurs

les transitions correspondant au message à coder) grâce à un algorithme de Viterbi. L'extraction du message est faite en décodant les données reçues y', toujours par Viterbi, mais cette fois en considérant l'ensemble des chemins possibles (le treillis de la figure 3.7(b)).

Miller *et al.* conseillent dans leur article d'utiliser un treillis dont tous les sommets de $t+1$ sont accessibles depuis n'importe quel sommet de t (c'est-à-dire 2^r arcs par sommet et par bit). Or, le pouvoir de correction du treillis est très réduit dans ce cas. Ils compensent en valuant les arcs par de très longues séquences pseudo-aléatoires, assurant ainsi une distance minimale correcte.

Le codage par syndrome, introduit par Pradhan et Ramchandran [PR00], est vu par Chou *et al.* [CPR99, CPGR00, CPR01] comme une façon de construire un code structuré utilisable dans le cadre de canaux avec information adjacente, et donc dans le tatouage.

Tout code correcteur binaire peut être vu comme l'ajout de bits de parité aux bits du message que l'on souhaite coder [CMB02]. La notation de l'équation 3.9 peut donc s'écrire

$$\begin{array}{rcl} \{0,1\}^k & \longrightarrow & \mathcal{U} \subset \{0,1\}^{k+(n-k)} \\ f : \mathsf{m} & \longmapsto & \mathsf{u} = \mathsf{m} \bullet \mathsf{p} \end{array} \tag{3.19}$$

où $\mathsf{a} \bullet \mathsf{b}$ représente la concaténation des vecteurs a et b. Le syndrome d'un mot est la différence entre les bits de parité de ce mot et les bits de parité normalement obtenu

par codage du message :

$$\text{syn}(\mathsf{u}' = \mathsf{m} \bullet \mathsf{p}') = \text{les } n-k \text{ derniers bits de } f(\mathsf{m}) \oplus \mathsf{u}'. \tag{3.20}$$

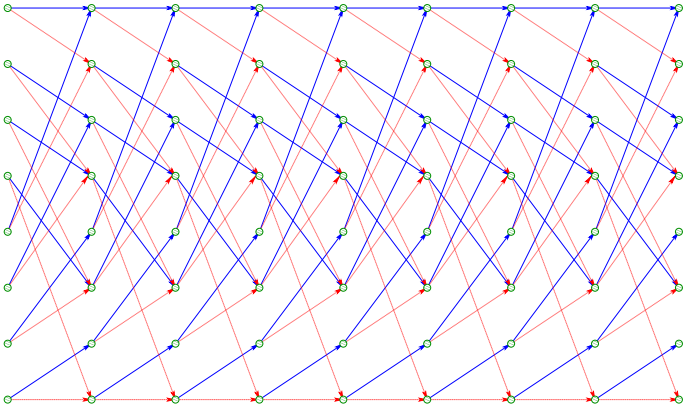
Un syndrome différent de $\mathsf{0}$ indique donc que le mot de code reçu comporte une ou plusieurs erreurs, mais n'informe pas sur la manière de corriger.

Le codage par syndrome consiste à envoyer le message *via* le syndrome du mot transmis. Ce type de codage a la propriété d'être structuré, comme l'exige le schéma de Costa. En effet, de nombreux mots peuvent donner le même syndrome et donc correspondre au même message transmis. Ainsi, si l'on souhaite coder le message $\mathsf{0}$, tous les mots de la forme $f(\mathsf{x})$ avec $\mathsf{x} \in \{0,1\}^k$ sont possibles. Afin de trouver le mot de code le plus proche, il faut décoder le signal hôte en limitant les chemins du treillis aux différents mots de code correspondant au message à transmettre. Cela est fait par un algorithme de Viterbi appliqué à un treillis dont les valeurs des arcs sont changés afin de correspondre au bon syndrome. Ce principe est très similaire à celui des codes de Miller.

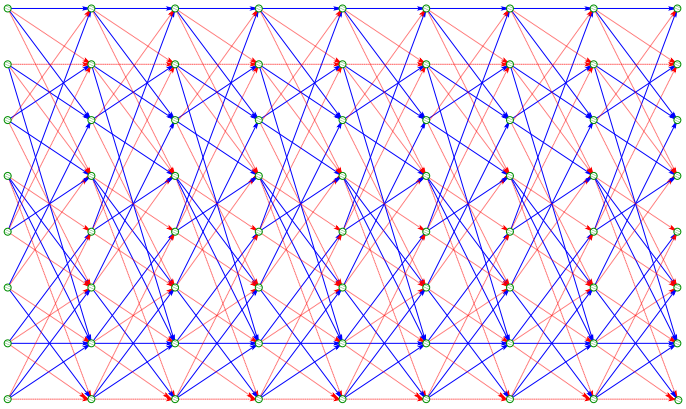
Conclusion

Transmettre un message *via* un document pouvant être attaqué est assimilable à un problème de communication. L'analogie entre le tatouage et le codage canal a été faite assez rapidement dans la littérature. En assimilant le signal hôte à un canal bruité par des attaques, le codage canal permet une modélisation théorique du tatouage et donne les formules de capacité ou de probabilités d'erreur associées. De plus, le codage canal apporte des outils tels que les codes correcteurs, permettant d'obtenir de forts gains de performances (baisse de probabilité d'erreur ou augmentation de la robustesse). Toutefois, même si de nombreux auteurs proposent d'utiliser des codes correcteurs pour améliorer les performances de leur schéma, peu ont véritablement considéré le tatouage comme de la communication et définit leur schéma en conséquence. On voit plutôt des techniques de communication se greffer plus ou moins heureusement sur des schémas de tatouage existants

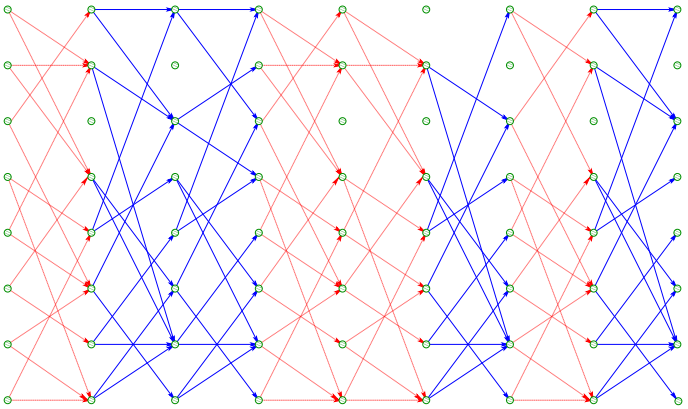
Enfin, les canaux avec information adjacente disponible à l'encodage trouvent avec le tatouage une application. Les travaux théoriques sur ce type de canaux ont montré qu'il était possible d'obtenir des performances bien supérieures aux limites obtenues en considérant le tatouage comme un problème de canal bruité classique. Néanmoins, le schéma théorique amenant à la démonstration de ces performances ne peut être utilisé directement, et les dérivations pratiques (notamment celles basées sur la quantification) restent sous-optimales.



(a) Treillis convolutif classique (un arc par sommet et par bit) à 8 états et 8 symboles



(b) Treillis avec deux arcs par sommet et par bit



(c) Treillis permettant de coder $\mathsf{m} = \{10011010\}$

FIG. 3.7 – Construction d'un dictionnaire structuré par un treillis convolutif selon [MDC02] (arcs bleus pour le bit 0 et arcs rouges pointillés pour le bit 1)

Chapitre 4

Adapter la marque au signal hôte

Bien que nous ayons vu dans le chapitre précédent que le tatouage est assimilable à la transmission d'un message sur un canal bruité, une différence importante entre le codage canal et le tatouage est que ce dernier s'inscrit au sein d'un document, dont les qualités perceptuelles ne doivent pas être endommagées. Coder le message sans tenir compte du signal hôte, en respectant uniquement la contrainte de distorsion maximale, peut donc amener à un signal marqué fortement dégradé.

Les premières tentatives pour prendre en compte l'impact perceptuel dans le tatouage ont été empiriques et basées sur des observations. L'introduction de mesures perceptuelles (voir la section 2.3) a permis ensuite une adaptation plus fine. Dernièrement, ces techniques sont remises en cause par l'utilisation de la théorie des jeux : la prise en compte du comportement de l'attaquant permet de redéfinir la stratégie d'allocation de la marque.

4.1 Techniques empiriques

La mesure de fidélité la plus courante entre deux signaux est le PSNR (équation 2.6). Elle ne fait pas intervenir d'aspect perceptuel, et donc tous les échantillons sont de même importance dans ce calcul. Si l'on reste dans cette optique, il est logique de construire un signal de marque dont l'énergie est répartie uniformément sur chaque échantillon : $\forall i \in \{1, 2, \ldots, n\}$, $\sigma^2_{W_i} = \mathrm{eqm}(\mathsf{x}, \mathsf{y})$.

Une des caractéristiques du système visuel humain, vu dans la section 2.3.1, est que l'œil est plus sensible aux modifications touchant les basses fréquences (zones uniformes) plutôt que les hautes fréquences (contours et zones très texturées). Une idée est donc de privilégier ces dernières lors de la répartition de l'énergie de la marque. Néanmoins, les hautes fréquences sont sujettes à de fortes dégradations lors d'attaques comme les compressions JPEG ou MP3, dégradations qui peuvent supprimer la marque. Il faut donc trouver un compromis entre invisibilité et robustesse.

L'idée la plus évidente est de marquer uniquement dans les moyennes fréquences. Ainsi, Csurka *et al.* [CDRP99b] utilisent une transformée de Fourier et sélectionnent les coefficients correspondant aux moyennes fréquences du spectre. Une autre approche

est donnée par Piva *et al.* [PBBC97, PBBC98], utilisant le domaine DCT. Les premiers coefficients de l'organisation en zig-zag sont évités (les plus basses fréquences) et les autres sont tatoués en utilisant

$$\begin{aligned} y_i &= x_i + w_i && (4.1)\\ &= x_i + \alpha|x_i| \times \overline{w}_i, && (4.2) \end{aligned}$$

où α, dont la valeur est commune à tous les échantillons de x, sert à régler le compromis robustesse/invisibilité du tatouage, et $\overline{w}_i$ est le codage du message à transmettre issu par exemple d'un étalement de spectre, suivant une loi Normale $\mathcal{N}(0,1)$. Cette forme d'insertion, où la force de la marque dépend directement de la valeur absolue du coefficient hôte, est justifiée par le fait que le seuil de visibilité dépend de l'amplitude du signal hôte[1].

4.2 Adaptation perceptuelle

La méthode de Piva vue au-dessus tente de prendre en compte un aspect perceptuel lors de l'insertion de la marque, mais de façon assez simpliste. Plus généralement, comme le remarquent Cox *et al.* [CMB02], il y a essentiellement deux possibilités de considérer l'impact perceptuel du tatouage. La première[2] est de choisir une distorsion s'appuyant sur un critère psycho-visuel ou psycho-acoustique, comme par exemple une MSE pondérée comme décrit dans la section 2.3.2, au lieu du classique PSNR et de ses dérivés. Dans ce premier cas, l'insertion est du type $y_i = x_i + \alpha\overline{w}_i$ avec $\overline{W} \sim \mathcal{N}(0,1)$, et α est réglé afin de respecté une distorsion D_{xy} maximale. Ainsi, une image très texturée pourra être marquée plus fortement (paramètre α important) qu'une image avec de nombreuses régions uniformes. Cette technique permet d'adapter la force globale de la marque signal par signal. Cela est illustré par la figure 4.1 : l'image *Baboon* est très texturée, à l'opposé de l'image *Rose*. Avec la même contrainte de distorsion[3] $D_{\mathsf{xy}} = 20$, l'image *Baboon* peut recevoir une marque donc l'énergie est quatre fois supérieure à celle que peut recevoir *Rose*, c'est-à-dire $\sigma_W^2 = \alpha^2 \simeq 414$ pour *Baboon* et seulement $\sigma_W^2 \simeq 103$ pour *Rose*.

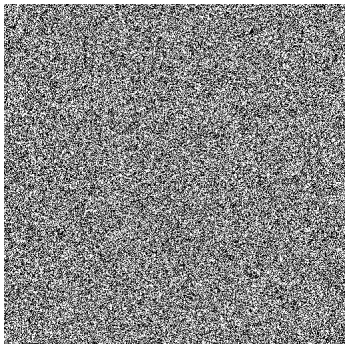
Cette méthode d'ajustement conduit à une marque répartie uniformément sur l'ensemble de l'image (on remarque bien un bruit blanc sur les figures 4.1(b) et 4.1(e)). Il serait intéressant d'adapter la marque zone par zone, à l'intérieur de l'image, comme le fait Piva avec son $w_i = \alpha|x_i| \times \overline{w}_i$. Avec un critère perceptuel, on aurait beaucoup d'énergie sur les contours et peu dans les basses fréquences. Cette adaptation est appelée *perceptual shaping* par Cox, et l'insertion est du type $y_i = x_i + \alpha_i \times \overline{w}_i$ avec α_i dépendant de l'importance perceptuel de l'échantillon x_i, noté ici φ_i. On peut avoir un simple $\alpha_i \propto \varphi_i^{-1}$ ou une combinaison avec un seuil de type JND (section 2.3.3) :

[1] *The perceptibility threshold of a sinusoidal grating depends on the amplitude of the iso-frequency signal to which it is superimposed* [BBCP98].

[2] Désignée par *perceptually limited embedding* par Cox.

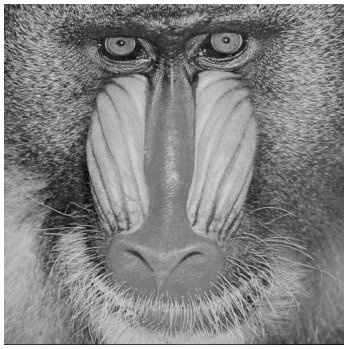
[3] Contrainte calculée en utilisant une erreur quadratique moyenne pondérée (voir l'équation 2.8) avec une pondération φ_i de Watson [Wat93].

(a) Image originale x de taille 512×512 : *Rose*

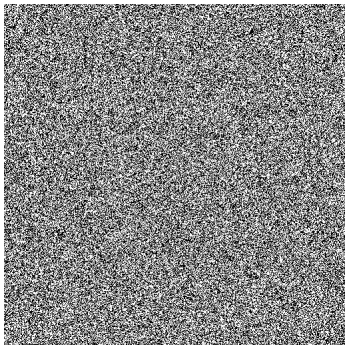


(b) Marque w, avec $\alpha = 10,127$

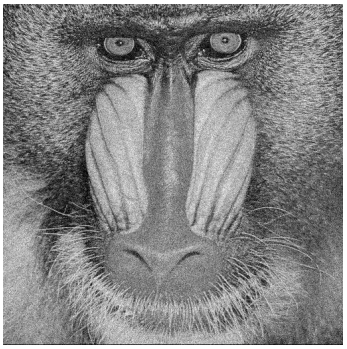
(c) Image marquée $\mathsf{y} = \mathsf{x} + \mathsf{w}$



(d) Image originale x de taille 512×512 : *Baboon*



(e) Marque w, avec $\alpha = 20,349$



(f) Image marquée $\mathsf{y} = \mathsf{x} + \mathsf{w}$

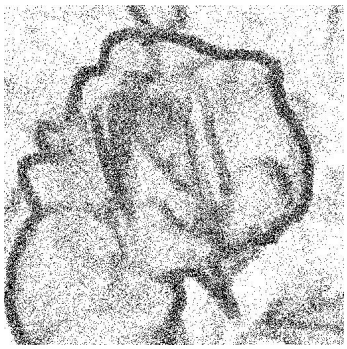
FIG. 4.1 – Insertion de type $y_i = x_i + \alpha \times s_i$ avec une contrainte de distorsion $D_{\mathsf{xy}} = \mathrm{weqm}(\mathsf{x}, \mathsf{y}) = 20$ identique pour les deux images

$\alpha_i = \min\left(\mathrm{cst} \times \varphi_i^{-1}, \mathrm{jnd}_{\mathsf{x}}(i)\right)$ afin de ne pas dépasser le seuil de visibilité/audition. Cette seconde technique est illustrée par la figure 4.2. Comme précédemment, la même contrainte de distorsion $D_{\mathsf{xy}} = 20$ est imposée aux marques pour les images *Baboon* et *Rose*. On remarque bien sur les images 4.2(b) et 4.2(e) que la marque est adaptée au signal hôte et est fortement concentrée sur les contours. La très grande majorité des techniques prenant en compte l'impact perceptuel du tatouage utilise une adaptation de ce type [DN00, JCL02].

4.3 Résolution par théorie du jeu

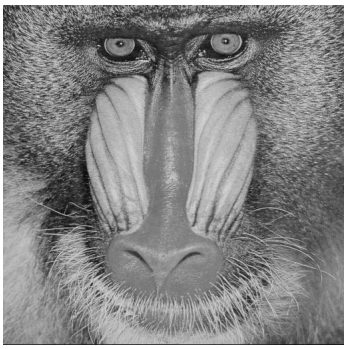
Adapter la marque en suivant son intuition ou en s'appuyant sur des mesures perceptuelles comme dans les techniques précédentes met de côté toute une partie de la

(a) Image originale x de taille 512×512 : *Rose*

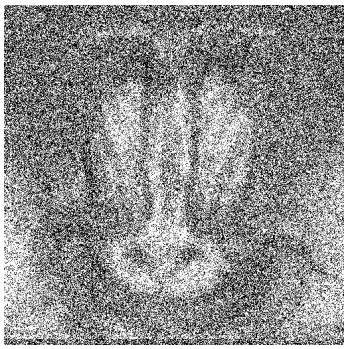


(b) Marque w, avec $\alpha_i = 4,47/\varphi_i$

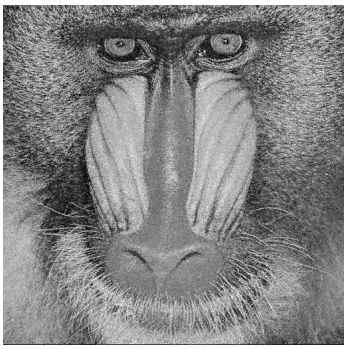
(c) Image marquée $\mathsf{y} = \mathsf{x} + \mathsf{w}$



(d) Image originale x de taille 512×512 : *Baboon*



(e) Marque w, avec $\alpha_i = 4,47/\varphi_i$



(f) Image marquée $\mathsf{y} = \mathsf{x} + \mathsf{w}$

FIG. 4.2 – Insertion de type $y_i = x_i + \alpha_i \times \overline{w}_i$ avec une contrainte de distorsion $D_{\mathsf{xy}} = \mathrm{weqm}(\mathsf{x}, \mathsf{y}) = 20$

chaîne de transmission. Comme le rappelle la figure 1.1 (page 16), la transmission du message passe par trois étapes : insertion, transmission (attaques) et extraction. Afin de garantir la meilleure robustesse, il faut considérer ces trois phases pour adapter la marque.

Cette remarque a été faite dès 1997 par Cox *et al.* [CKLS97]. Si l'on considère la dernière technique de la section précédente, consistant à prendre une marque dépendante d'un facteur perceptuel local en chacun des échantillons à marquer, la marque est principalement concentrée sur les composantes les moins significatives d'un point de vue perceptuel (contours de la figure 4.2(b)), afin de maximiser l'énergie globale de la marque tout en respectant la contrainte de distorsion. Or, ce sont ces composantes qui seront attaquées en priorité par les traitements les plus usuels, comme la compressions JPEG ou MP3. En effet, si l'on veut supprimer ou modifier le plus d'éléments d'un signal, tout en minimisant l'impact perceptuel, il est logique de s'attaquer en priorité aux composantes les moins significatives. Ainsi, si le marqueur (la phase d'insertion) et l'attaquant agissent isolément, ils vont se concentrer sur les mêmes éléments. Au final, la marque risque d'être très dégradée. Ce raisonnement aboutit à une conclusion assez tranchée de la part de Cox : *the watermark should not be placed in perceptually insignificant regions of the image (or its spectrum) since many common signal and geometric processes affect these components.* Elle aura néanmoins de nombreux adeptes par la suite [WPD99, SC02, HWS02].

De façon plus générale, la relation entre la phase d'insertion et les techniques d'attaque peut être vue comme un jeu entre un attaquant et un défenseur [OME98]. Considérons une mesure de performance du schéma, telle que la probabilité d'erreur $\mathbf{P}_\mathrm{e}$ ou la capacité du canal de tatouage. L'attaquant veut minimiser la performance en respectant une contrainte de distorsion maximale $D_\mathrm{a}^{\max}$, tandis que le défenseur veut maximiser la performance avec une contrainte de distorsion d'insertion $D_\mathsf{xy}^{\max}$.

La première tentative de résolution du jeu fut donnée par Su *et al.* [SG99]. Ils considèrent des distorsions de type EQM (pas de pondération perceptuelle) et analysent plusieurs couples attaque/défense. La première attaque est tout simplement l'ajout de bruit gaussien (et donc un canal AWGN comme décrit dans la section 3.1.2). La meilleure stratégie à adopter dans ce cas du point de vue du défenseur est d'utiliser une marque de spectre proportionnellement similaire à celui du signal hôte. Cette condition est baptisée PSC pour *power-spectrum condition* et est résumée par : *the watermark should look like the original.* On peut remarquer que la technique de Piva de type $w_i = \alpha|x_i| \times \overline{w}_i$ respecte le PSC, et que d'autres papiers aboutissent à une forme de marque équivalente [SKH02].

Le simple ajout de bruit gaussien étant sous-optimal, ils définissent une attaque optimale dans un papier plus récent [SEG01a]. Elle se compose de l'ajout de bruit gaussien dans les cas où la distorsion d'attaque possible est faible, et de la suppression de l'échantillon considéré pour les distorsions les plus fortes (que l'on peut modéliser par un canal SAWGN, section 3.1.3). Mais dans cette configuration, une marque PSC n'est pas optimale, et ils proposent une méthode itérative numérique afin de trouver la solution du jeu.

Analytiquement, le jeu se formule par une optimisation de type max-min [CL00, CL01, Mou01, MI01a]. Considérons l'ensemble $\mathcal{E}(D_{\mathsf{xy}}^{\max})$ des répartitions possibles de la marque telles que celles-ci respectent la contrainte de distorsion maximale $D_{\mathsf{xy}}^{\max}$, et l'ensemble $\mathcal{A}(D_{\mathsf{a}}^{\max})$ de toutes les attaques possibles respectant la contrainte $D_{\mathsf{a}}^{\max}$. La performance optimale est donnée par

$$p^{\star} = \max_{\mathsf{e}\in\mathcal{E}(D_{\mathsf{xy}}^{\max})} \min_{\mathsf{a}\in\mathcal{A}(D_{\mathsf{a}}^{\max})} \mathrm{perf}(\mathsf{e},\mathsf{a}), \tag{4.3}$$

où perf(e,a) est la performance de la transmission du schéma de tatouage (capacité, probabilité $1-\mathbf{P}_{\mathrm{e}}$, ...) avec une énergie répartie suivant e et après une attaque a. Les papiers traitant de la résolution de l'équation 4.3 diffèrent par les gammes d'attaques possibles, la distribution supposée du signal hôte et les mesures de distorsion. Ainsi, Cohen et Lapidoth [CL02] considérent un signal gaussien et uniformément distribué, avec une distorsion de type EQM. Ils calculent la capacité du canal de tatouage soumis à une attaque de type SAWGN (ajout d'un bruit gaussien z et facteur d'échelle $\bar{\bar{\gamma}}$). Le premier résultat est obtenu en utilisant une mesure de distorsion d'attaque calculée entre le signal marqué et le signal attaqué. Ils trouvent alors

$$C = \frac{1}{2}\log_2\left[1+\frac{D_{\mathsf{xy}}}{D_{\mathsf{yy}'}}\right], \tag{4.4}$$

avec

$$D_{\mathsf{xy}} = \sum_{i=1}^{n}(x_i-y_i)^2 \tag{4.5}$$

$$\text{et } D_{\mathsf{yy}'} = \sum_{i=1}^{n}(y_i-y'_i)^2. \tag{4.6}$$

On reconnaît la capacité de l'équation 3.17. En utilisant une mesure de distorsion d'attaque calculée entre le signal d'origine x et le signal marqué puis attaqué y', la formule de capacité est alors

$$C = \frac{1}{2}\log_2\left[1+\frac{D_{\mathsf{xy}}}{D_{\mathsf{xy}'}-D_{\mathsf{xy}}}\left[1-\frac{D_{\mathsf{xy}'}}{\sigma_X^2}\right]\right], \tag{4.7}$$

où σ_X^2 est la variance du signal hôte. Les stratégies optimales pour l'attaquant et le défenseur prennent les formes suivantes :

$$\mathsf{y} = \bar{\gamma}\,(\mathsf{x}+\mathsf{w}) \tag{4.8}$$

$$\mathsf{y}' = \bar{\bar{\gamma}}\,(\mathsf{y}+\mathsf{z})\,. \tag{4.9}$$

Les paramètres sont donnés par

$$\bar{\gamma} = \frac{\sigma_X^2 - D_{\mathsf{xy}}}{\sigma_X^2} \tag{4.10}$$

$$\bar{\bar{\gamma}} = \frac{\sigma_X^2 - D_{\mathsf{xy}'}}{\sigma_X^2 - D_{\mathsf{xy}}} \tag{4.11}$$

$$\sigma_Z^2 = (D_{\mathsf{xy}'} - D_{\mathsf{xy}})\frac{\sigma_X^2 - D_{\mathsf{xy}'}}{\sigma_X^2 - D_{\mathsf{xy}}}. \tag{4.12}$$

On voit depuis l'équation 4.8 que le simple tatouage additif n'est pas le plus efficace. On remarque également que le facteur $\bar{\gamma}$ correspond à l'expression d'un filtre de Wiener dont le but serait de réduire l'impact de la marque w. De la même façon, le facteur $\bar{\bar{\gamma}}$ est en fait l'expression du filtre de Wiener atténuant la marque et le bruit ajouté $\mathsf{z}/\bar{\bar{\gamma}}$. L'attaque consiste à ajouter du bruit et à le filtrer.

Des travaux similaires ont été proposés par Moulin [Mou01, MM03], étendus à des signaux non identiquement distribués. Il considère un signal hôte non i.i.d. séparé en canaux gaussiens parallèles (déjà évoqués dans la section 2.2). La capacité totale est alors la somme des capacités atteignables sur chacun des k canaux :

$$C = \max_{d_{\mathsf{xy}}} \min_{d_{\mathsf{xy}'}} \sum_{i=1}^{k} r_i \times \Gamma(\sigma_i^2, d_{\mathsf{xy}}(i), d_{\mathsf{xy}'}(i)), \tag{4.13}$$

où σ_i^2 est la variance du $i^{\text{ème}}$ canal, r_i le rapport entre nombre d'échantillons du $i^{\text{ème}}$ canal et nombre total d'échantillons, $d_{\mathsf{xy}}(i)$ la distorsion d'insertion sur ce canal, $d_{\mathsf{xy}'}(i)$ la distorsion d'attaque. La fonction $\Gamma()$ mesure la capacité atteignable avec les caractéristiques passées en paramètres (voir l'équation 4.7). On retrouve le jeu entre attaquant et défenseur de l'équation 4.3 : l'attaquant recherche une répartition $d_{\mathsf{xy}'}$ de sa distorsion afin de minimiser capacité du canal, tout en respectant

$$\sum_{i=1}^{k} r_i \times d_{\mathsf{xy}'}(i) \leq D_{\mathsf{xy}'}^{\max}, \tag{4.14}$$

et le défenseur souhaite maximiser cette capacité avec une contrainte de distorsion de la même forme. L'auteur précise que les résultats peuvent être étendues à des distorsions de type wEQM (avec pondération perceptuelle, voir la section 2.3.2) en multipliant chaque échantillon x_i du signal hôte par $\sqrt{\varphi_i}$, où φ_i est la pondération correspondante [MM01].

Conclusion

Même si le codage canal apporte des solutions pour coder le message à transmettre, il ne tient pas compte des contraintes d'invisibilité du tatouage, ni des interactions entre marqueur et attaquant. La répartition de la marque au sein du document hôte est un facteur important sur l'impact de la marque et sa robustesse.

Si les premières idées de répartitions différentes de l'uniformité furent dictées par l'intuition, la prise en compte de facteurs perceptuels a permis depuis de résoudre le problème de l'invisibilité, mais pas celui de la robustesse vis-à-vis des attaques volontaires. Voir les interactions entre marqueur et attaquant comme un jeu autorise l'utilisation d'optimisation de type max-min : en considérant une mesure de performance

commune aux deux parties et des contraintes de distorsion, il est possible de définir une attaque optimale et également la répartition optimale de la marque permettant d'obtenir la meilleure performance.

Deuxième partie

Étalement de spectre pour le tatouage

Introduction

La première partie de ce manuscrit a donné un aperçu des tendances en terme de tatouage. L'une des principales conclusions est que le tatouage est un problème de communication. Il est à ce titre légitime d'utiliser les notions et les outils développés dans ce domaine. Un point fait néanmoins la différence entre tatouage et simple communication : le document hôte. La contrainte imposant que l'information présente lors de l'encodage du message à transmettre (considérée comme du bruit) doit rester utilisable malgré l'ajout d'une marque et après attaque est inédit en terme de communication. La transmission d'un message doit donc à la fois jouer avec les notions de codage (modulation, codes correcteurs, ...) mais aussi avec des distorsions, des aspects perceptuels et déjouer des stratégies d'attaque. On sépare ce problème en deux aspects. Il faut tout d'abord définir un canal de tatouage. Cette définition doit prendre en compte des contraintes de distorsion afin que la marque soit quasi-imperceptible. Il faut pour cela adapter l'énergie de la marque en fonction du signal hôte (chapitre 4 de la première partie), mais aussi en fonction des attaques que le document marqué sera susceptible de rencontrer. L'autre aspect concerne le codage du message. Les travaux de Costa [Cos83], présentés dans le chapitre 3, trouvent une excellente application dans le tatouage, et montrent qu'il est possible d'obtenir de bien meilleures performances avec un codage adapté.

Plusieurs travaux ont tenté de définir un canal de tatouage prenant en compte une distorsion et d'éventuelles attaques. Les plus performants modélisent l'interaction entre l'insertion et les attaque par un jeu. Les articles de Cohen et Lapidoth [CL00, CL01], présentés dans la section 4.3 de la partie 1, ont permis de définir la forme optimale de la marque (ajout d'une marque puis filtrage de Wiener) et la forme de l'attaque (bruit et filtrage) dans le cas de signaux identiquement distribués et gaussiens. Une limite de capacité en est déduite. Or, les signaux multimedia généralement rencontrés suivent rarement une simple loi Normale (voir la section 2.2 de la page 31).

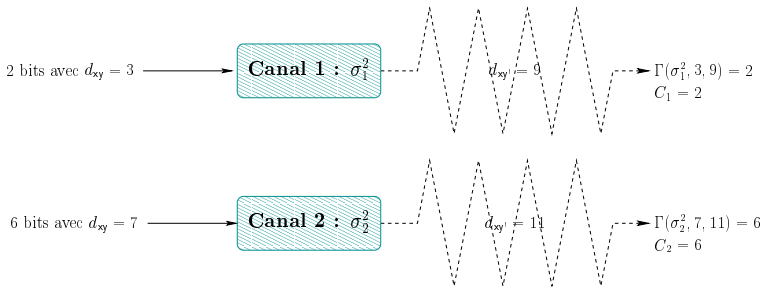
Leurs derniers travaux [CL02] et ceux de Moulin [Mou01, MM03] étendent cette étude aux signaux non i.i.d. en subdivisant le signal hôte en canaux parallèles gaussiens. Les échantillons du signal qui suivent une loi similaire sont regroupés en sous-canaux[4]. Une optimisation de type max-min permet de définir la forme de l'attaque optimale et

[4]Chaque échantillon est supposé suivre une loi $\mathcal{N}(0, \sigma^2_{X_i})$. La variance est calculée en utilisant une fenêtre centrée sur l'échantillon considéré. Toutes ces variances sont quantifiées, donnant ainsi des groupes d'échantillons aux propriétés statistiques proches.

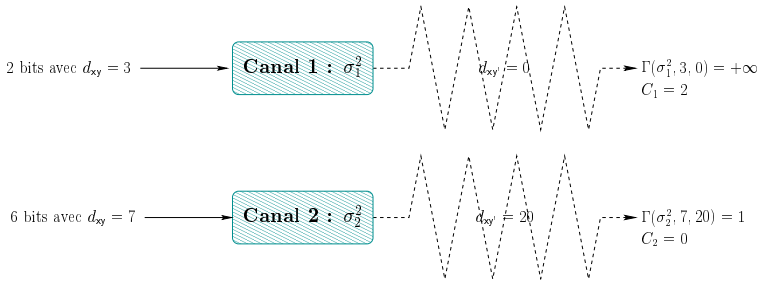
la défense (la répartition de l'énergie de la marque ajoutée) correspondante. La capacité totale est la somme des capacités obtenues sur chacun des canaux. Néanmoins, le fait d'optimiser sur des canaux indépendants pose des problèmes si l'attaquant ne choisit pas l'attaque optimale telle que définie par le jeu. Un exemple fictif est donné par la figure 3, en considérant deux canaux. La distorsion d'insertion maximale est de 10, et la distorsion d'attaque de 20. La résolution du jeu nous donne la meilleure répartition de l'attaque : $d_{\mathsf{xy}'}(1) = 9$ sur le premier canal et $d_{\mathsf{xy}'}(2) = 11$ sur le second. La meilleure stratégie de défense est alors de prendre $d_{\mathsf{xy}}(1) = 3$ et $d_{\mathsf{xy}}(2) = 7$. La capacité totale atteignable dans ces conditions est alors de 8 bits. Deux bits sont donc transmis sur le premier canal et six sur le second. Or, l'attaquant a intérêt à changer de stratégie. Ainsi, sur la figure 3(b), il concentre son attaque sur le second canal. Les six bits transmis ne seront probablement pas extraits correctement, faisant chuter la capacité totale (seuls les deux bits du canal 1 seront corrects). Pourtant, cette stratégie d'attaque n'est pas optimale : la capacité totale est potentiellement infinie car un canal n'est pas attaqué. Mais comme le cas n'a pas été prévu, seuls deux bits ont été transmis. Une fois la stratégie d'insertion fixée, l'attaquant a donc tout intérêt à choisir une stratégie d'attaque autre que celle normalement prévue. Le jeu n'aboutit pas à un équilibre. Un autre point sensible de ce type de canaux est que le receveur doit pouvoir reconstruire les sous-canaux à partir du signal marqué et attaqué (une attaque perturbant la reconstruction en modifiant les propriétés statistiques des échantillons peut fortement perturber la transmission). Il doit également connaître le nombre de bits transmis sur chacun des canaux.

Afin d'éviter ses écueils, une solution est de transmettre l'ensemble des bits sur un canal global commun. De ce fait, une attaque plus faible que prévue ne pourra donner que de meilleurs résultats. Certes, si k bits sont insérés, il sera bien sûr impossible d'en extraire plus de k, mais cela pourra par exemple se répercuter par un gain en terme de probabilité d'erreur. Les problèmes concernant la subdivision du signal hôte en sous-signaux sont aussi évités. Une solution est l'étalement de spectre, dont le principe a déjà été évoqué dans la section 1.2.1 de la première partie. La porteuse correspondant à chaque symbole transmis est répartie uniformément sur l'ensemble du signal hôte. Si l'attaque se concentre uniquement sur une partie du document (pour un même niveau de distorsion totale), cela va réduire équitablement la performance d'extraction pour chacun des bits. Mais en contrepartie, cette perte sera compensée car le reste du document ne sera pas attaqué : la performance de la transmission ne pourra être inférieure à celle prévue par une optimisation de type max-min si l'on obtient un équilibre.

Cette partie porte donc sur l'optimisation d'un schéma de tatouage basé sur l'étalement de spectre. Le premier chapitre va introduire les notations utilisées, puis nous définirons la technique d'extraction du message inséré en considérant des attaques de type SAWGN. Cela aboutira à une mesure de performance. Elle sera utilisée dans le chapitre suivant, au sein d'un jeu entre un attaquant et un défenseur. On verra ainsi la forme de l'attaque optimale quelle que soit la stratégie d'insertion choisie, puis la répartition de l'énergie de la marque à privilégier afin d'y résister au mieux. Le



(a) L'attaque subie est celle prévue : les 8 bits sont correctement transmis



(b) L'attaque se concentre sur un canal : la capacité totale est de 2 bits

FIG. 3 – Les canaux parallèles gaussiens face à deux répartitions d'attaque

chapitre 3 reprend ce jeu en considérant une insertion additive suivie d'un filtrage, ce qui donne une nouvelle forme de répartition. Enfin le dernier chapitre démontre expérimentalement l'efficacité de l'attaque ainsi définie et les performance de notre technique de défense.

Chapitre 1

Présentation du problème

Le tatouage par étalement de spectre permet de coder le message à transmettre sur l'ensemble des échantillons du signal hôte. Une attaque éventuelle aura le même impact sur chacun des symboles du message, contrairement à la technique des canaux gaussiens parallèles utilisée par Moulin.

Nous verrons d'abord dans la suite de ce premier chapitre un rappel du principe du tatouage par étalement de spectre avec l'introduction des notations utilisées. Puis nous introduirons une modélisation des attaques liées à la diffusion du document marqué en s'appuyant sur les canaux SAWGN. Enfin, en considérant ces hypothèses d'insertion et d'attaque, nous exposerons la méthode d'extraction optimale.

1.1 Insertion par étalement de spectre

L'étalement de spectre [PWM82] est une technique de communication permettant de transmettre un message sur un canal très bruité. Il est par exemple utilisé dans les réseaux sans fil ou encore dans les systèmes de positionnement GPS [Dix94]. À chaque symbole possible et à chaque position du symbole dans le message à transmettre est associé un vecteur porteur. La somme des ces vecteurs va constituer le signal à transmettre. L'extraction se fait en recherchant les vecteurs porteurs au sein du signal reçu. La robustesse de l'étalement de spectre en fait notre choix de prédilection pour le schéma de tatouage présenté ici.

Nous considérons un message à transmettre noté $\mathsf{b} \in \mathbb{R}^n$ modélisé par les v.a. B_j (avec $j \in \{1, 2, \ldots, n\}$) et d'énergie $\mathbb{E}\left[B_j^2\right] = 1$. Il peut être obtenu par codage direct des informations que l'on souhaite transmettre, ou bien par la sortie d'un système de correction d'erreurs. Le signal hôte est noté $\mathsf{x} = \{x_1, x_2, \ldots, x_m\}$. Comme vu dans la première partie, il existe de nombreux modèles statistiques adaptés aux différents signaux multimedia, selon la transformée fréquentielle dont ils sont issus ou selon le type de document qu'ils représentent. Néanmoins, il apparaît que la modélisation sous forme de lois Normales indépendantes pour chacun des échantillons de x permet la prise en compte de nombreux cas, comme les mixtures de gaussiennes pour les co-

efficients issus d'une transformée en ondelettes [CNB98, WV99], ou les canaux parallèles gaussiens [hKRM99, Mou01]. Nous supposons donc le signal hôte comme la réalisation d'un ensemble de variables aléatoires indépendantes $\mathsf{X} = \{X_1, X_2, \ldots, X_m\}$, avec $X_i \sim \mathcal{N}(0, \sigma^2_{X_i})$. Nous supposons que la valeur de n est très inférieure celle de m, ce qui est généralement le cas en pratique, la taille du message à transmettre dans les schémas de tatouage robuste étant tout au plus de quelques centaines de bits, alors que le signal hôte se compose souvent de plus de cent mille échantillons (un signal hôte composé de la luminance d'une image monochrome de taille 512×512 donne $m > 2,5 \times 10^5$ et un extrait sonore de qualité CD de 30 secondes donne $m \simeq 1,3 \times 10^6$).

Afin d'insérer n symboles binaires dans m échantillons, l'étalement de spectre utilise un ensemble de porteuses, regroupées ici au sein d'une matrice G de dimensions $n \times m$ et telle que $\mathsf{G} \in \mathbb{R}^{n.m}$, dont chacun des éléments est d'espérance nulle et d'énergie $\mathbb{E}\left[\mathsf{G}(i,j)^2\right] = 1$. Cette matrice est construite pseudo-aléatoirement en utilisant une clef $c \in \mathcal{C}$. Dans le contexte de cette étude, nous nous focalisons sur un schéma de tatouage symétrique. La clef c est donc connue lors de l'insertion et de l'extraction, mais reste secrète pour les autres protagonistes de la chaîne de communication (les éventuels attaquants dans notre cas). Le signal de marque w est défini par

$$\forall i \in \{1, 2, \ldots, m\}, \quad w_i = \frac{\sigma_{W_i}}{\sqrt{n}} \sum_{j=1}^{n} \mathsf{G}(i,j) \times b_j, \tag{1.1}$$

et le signal marqué est

$$\mathsf{y} = \mathsf{x} + \mathsf{w}. \tag{1.2}$$

La pondération $\sigma_{W_i}/\sqrt{n}$ permet d'adapter l'énergie de la marque échantillon par échantillon. Par le théorème de la limite centrale, comme chaque élément w_i est obtenu par la somme de n valeurs pseudo-aléatoires (avec n suffisamment grand), la marque peut être vue comme la réalisation d'un ensemble de variables aléatoires $\mathsf{W} = \{W_1, W_2, \ldots, W_m\}$ suivant une loi Normale : $W_i \sim \mathcal{N}(0, \sigma^2_{W_i})$. Et comme les signaux x et w sont indépendants, y est une réalisation de Y avec $Y_i \sim \mathcal{N}(0, \sigma^2_{Y_i} = \sigma^2_{X_i} + \sigma^2_{W_i})$.

1.2 Modélisation des attaques

Nous ne prenons pas en compte dans cette partie les attaques géométriques (c'est-à-dire désynchronisantes), c'est-à-dire introduisant un décalage entre les données marquées y et les données attaquées y'. De nombreux articles considèrent les attaques non désynchronisantes comme un ajout de bruit [SPR98, MO99, MO00, CL02]. Mais ce modèle ne permet pas de prendre en compte des traitements évolués tels que le filtrage, les techniques de compression avec perte ou encore le bruit corrélé. Les canaux de type SAWGN, déjà évoqués dans la section 3.1.3 de la première partie, apportent en plus du simple ajout de bruit un facteur de pondération. Cela permet par exemple de modéliser l'ajout d'un bruit corrélé au signal hôte, de prendre en compte un facteur d'échelle (changement du volume sonore d'un extrait audio, variations de contraste et de luminosité d'une image, ...) ou un filtrage linéaire. Les données reçues s'écrivent dans ce

cas

$$\forall i \in \{1,2,\ldots,m\},\ \ y'_i = \gamma_i \times y_i + z_i \tag{1.3}$$

$$= \gamma_i \left[x_i + \frac{\sigma_{W_i}}{\sqrt{n}} \sum_{j=1}^{n} \mathsf{G}(i,j) \times b_j \right] + z_i, \tag{1.4}$$

où γ_i est un facteur compris d'échelle, et z_i est un bruit gaussien blanc, réalisation de la variable aléatoire $Z_i \sim \mathcal{N}(0, \sigma^2_{Z_i})$. Une autre formulation possible aurait pu être $y'_i = \gamma_i(y_i + z_i)$, mais notre formule est plus générale. Elle permet en effet de prendre en compte des attaques consistant à annuler la marque et à ajouter du bruit ($\gamma_i = 0$ et $\sigma_{Z_i} > 0$).

1.3 Extraction du message

Comme nous l'avons vu dans la section 1.2.1 de la première partie, dans le cas du tatouage additif, l'extraction se fait généralement par corrélation entre les porteuses et le signal reçu. Nous allons utiliser ici un critère de maximun *a posteriori* (MAP) afin d'estimer chacun des symboles b_j introduits par le tatouage. Dans ce cas de figure, la valeur estimée du $j^{\text{ème}}$ bit est

$$\widehat{b_j} = \arg\max_b \{P_j(b)\} \tag{1.5}$$

$$\text{avec } P_j(b) = \Pr\left(B_j = b \,|\, \mathsf{Y}' = \mathsf{y}'\right), \tag{1.6}$$

où Y' est l'ensemble de m variables aléatoires modélisant le signal y' et B_j est la v.a. modélisant le $j^{\text{ème}}$ bit inséré. Par la loi de Bayes, nous avons

$$P_j(b) = \frac{\Pr\left(B_j = b,\, \mathsf{Y}' = \mathsf{y}'\right)}{\Pr\left(\mathsf{Y}' = \mathsf{y}'\right)}$$

$$= \frac{\Pr\left(\mathsf{Y}' = \mathsf{y}' \,|\, B_j = b\right) \times \Pr\left(B_j = b\right)}{\Pr\left(\mathsf{Y}' = \mathsf{y}'\right)}. \tag{1.7}$$

Par définition, le signal y' est parfaitement connu à la réception, et donc $\Pr(\mathsf{Y}' = \mathsf{y}')$ est une constante. De plus, comme $\Pr(B_j = b) = 1/2$ (car nous n'avons pas d'*a priori* sur la valeur des bits composant le message), alors

$$P_j(b) \propto \Pr\left(\mathsf{Y}' = \mathsf{y}' \,|\, B_j = b\right). \tag{1.8}$$

Comme les sites marqués sont indépendants, on peut exprimer $P_j(b)$ par un produit de probabilités :

$$P_j(b) \propto \prod_{i=1}^{m} \Pr\left(Y'_i = y'_i \,|\, B_j = b\right) \tag{1.9}$$

$$\propto \prod_{i=1}^{m} \Pr\left(Y'_i - \frac{\gamma_i \sigma_{W_i}}{\sqrt{n}} \mathsf{G}(i,j) \times B_j = y'_i - \frac{\gamma_i \sigma_{W_i}}{\sqrt{n}} \mathsf{G}(i,j) \times b \,|\, B_j = b\right).$$

Le conditionnement de cette probabilité étant indépendant du reste, il peut être supprimé. On peut donc reformuler par

$$P_j(b) \propto \prod_{i=1}^{m} \Pr\left(Y'_i - \frac{\gamma_i \sigma_{W_i}}{\sqrt{n}} \mathsf{G}(i,j) \times B_j = y'_i - \frac{\gamma_i \sigma_{W_i}}{\sqrt{n}} \mathsf{G}(i,j) \times b\right). \tag{1.10}$$

Par les hypothèses de départ sur les distributions de X et Z, ce produit est un produit de distributions gaussiennes de formes

$$Y'_i - \frac{\gamma_i \sigma_{W_i}}{\sqrt{n}} \mathsf{G}(i,j) \times B_j \sim \mathcal{N}\left(0, \sigma_i^2\right) \tag{1.11}$$

$$\text{avec } \sigma_i^2 = \gamma_i^2 \left(\sigma_{X_i}^2 + \frac{(n-1)\,\sigma_{W_i}^2}{n}\right) + \sigma_{Z_i}^2. \tag{1.12}$$

Le produit de probabilités s'exprime alors par

$$P_j(b) \propto \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi} \times \sigma_i} \exp\left[-\frac{\left(y'_i - \frac{\gamma_i \sigma_{W_i} \times \mathsf{G}(i,j) \times b}{\sqrt{n}}\right)^2}{2\sigma_i^2}\right], \tag{1.13}$$

que l'on peut écrire

$$P_j(b) \propto \exp\left(\frac{-\Lambda(b)}{2}\right) \tag{1.14}$$

$$\text{avec } \Lambda(b) = \sum_{i=1}^{m} \frac{\left(y'_i - \frac{\gamma_i \sigma_{W_i} \times \mathsf{G}(i,j) \times b}{\sqrt{n}}\right)^2}{\sigma_i^2}. \tag{1.15}$$

Après quelques manipulations décrites en détails dans l'annexe A.1, on peut mettre $\Lambda(b)$ sous la forme

$$\Lambda(b) = \frac{\left(b - \overline{b_j}\right)^2}{\sigma_{b_j}^2} + \Gamma, \tag{1.16}$$

où Γ est indépendant de b, et

$$\sigma_{b_j}^{-2} = \sum_{i=1}^{m} \frac{\gamma_i^2 \sigma_{W_i}^2}{n \times \sigma_i^2} \quad \text{et} \quad \overline{b_j} = \sigma_{b_j}^2 \sum_{i=1}^{m} \frac{\gamma_i \sigma_{W_i} \times \mathsf{G}(i,j) \times y'_i}{\sqrt{n} \times \sigma_i^2}. \tag{1.17}$$

En reprenant l'équation 1.5, on déduit alors

$$\begin{aligned} \widehat{b_j} &= \arg\max_b \{P_j(b)\} \\ &= \arg\min_b \{\Lambda(b)\} \\ &\propto \overline{b_j}. \end{aligned} \tag{1.18}$$

L'expression de $\overline{b_j}$ est donc l'estimateur optimal au sens du MAP. Sa performance peut être mesurée en terme de rapport signal-à-bruit E_b/N_0 du canal de tatouage, qui s'avère être un canal gaussien (voir la section 3.1.2 de la première partie), exprimé par

$$\frac{E_b}{N_0} = \frac{\mathbb{E}\left[\widehat{b_j}^2\right]}{\sigma^2_{b_j}} = \frac{1}{n}\sum_{i=1}^{m} \frac{\gamma_i^2 \sigma^2_{W_i}}{\gamma_i^2 \left(\sigma^2_{X_i} + \sigma^2_{W_i}(n-1)/n\right) + \sigma^2_{Z_i}}. \tag{1.19}$$

Le résultat obtenu est en fait très conforme à l'intuition. On retrouve au nominateur l'énergie de la marque, pondérée par le facteur d'échelle de l'attaque, et au dénominateur les différents bruits intervenant lors de la transmission du message : le bruit du signal hôte, le bruit gaussien de l'attaque et enfin le bruit des autres symboles. Le rapport signal-à-bruit d'un canal gaussien nous permet de calculer sa capacité et la probabilité d'erreur (voir la section 3.1.2 de la première partie). Minimiser la probabilité d'erreur est équivalent à maximiser E_b/N_0.

On remarque également que l'estimateur est un produit de corrélation pondérée entre le vecteur porteur du symbole à extraire et le signal reçu y'. Il est proche de celui donné par Wu *et al.* [WYL02] déja évoqué dans la première partie (section 1.2.1 de la page 20) : la corrélation entre le signal reçu y' et la marque w est divisée par la somme du bruit ajouté lors de la transmission. On note également que ce résultat ne correspond pas à l'estimation de la marque par filtrage de Wiener, contrairement à ce qui a déjà été proposé [VDPP01].

1.4 Capacité du canal

Comme l'indique la formule 1.19, les n symboles transmis se partagent un rapport signal-à-bruit global. Moins on transmet de symboles et plus le rapport signal-à-bruit du canal sera important. On peut s'interroger sur la capacité d'un tel canal. Considérons une transmission de m symboles sur un canal gaussien caractérisé par le rapport signal-à-bruit E_b/N_0. La capacité totale maximale, selon ce qui a été vu dans la section 3.1.2 de la première partie, est donnée par

$$\mathcal{C}^{\max} = mC = \frac{m}{2}\log_2\left[1 + \frac{E_b}{N_0}\right]. \tag{1.20}$$

En utilisant l'étalement de spectre, nous nous limitons à n symboles. L'énergie disponible par symbole est multipliée par un facteur m/n. Par contre, le bruit ajouté reste le même. Malgré le fait que le signal hôte représente un bruit d'énergie supérieure à l'énergie de la marque, cela nous permet d'atteindre un rapport signal-à-bruit potentiellement important. La capacité totale est alors donnée par

$$\mathcal{C}^{\text{st}} = nC^{\text{st}} = \frac{n}{2}\log_2\left[1 + \frac{m \times E_b}{n \times N_0}\right]. \tag{1.21}$$

Pour une valeur E_b/N_0 faible, la formule de la capacité par échantillon peut être approximée par

$$C = \frac{1}{2}\log_2\left[1 + \frac{E_b}{N_0}\right] \simeq \frac{E_b}{2\ln 2 \times N_0}. \tag{1.22}$$

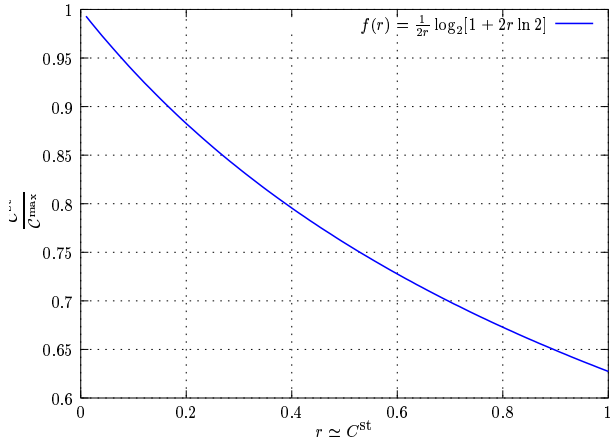


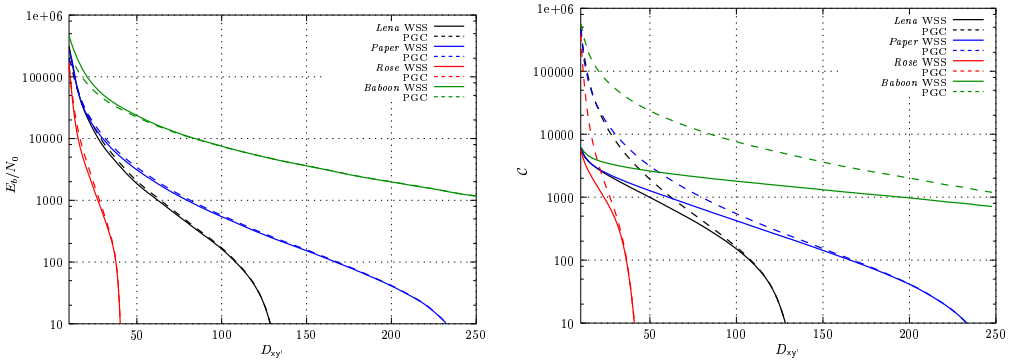
FIG. 1.1 – Capacité totale en utilisant l'étalement de spectre

On écrit alors l'équation 1.21 sous la forme

$$\begin{aligned} \mathcal{C}^{st} &= \frac{n}{2}\log_2\left[1+2C^{st}\ln 2\right] \\ &\simeq \mathcal{C}^{max}\times f(r) \text{ avec } f(r)=\frac{1}{2r}\log_2\left[1+2r\ln 2\right]. \end{aligned} \tag{1.23}$$

La fonction $f()$ représente le rapport entre la capacité totale maximale d'un canal gaussien classique (ou de canaux parallèles gaussiens) et la capacité qu'il est possible d'atteindre en utilisant l'étalement de spectre, et r est le rapport entre le nombre de bits utiles et le nombre de bits n (c'est-à-dire le rendement du code utilisé). La figure 1.1 montre cette fonction. On voit que pour obtenir la meilleure capacité, le rendement doit être le plus faible possible. Cette capacité est maximale ($\mathcal{C}^{st} = \mathcal{C}^{max}$) quand le rendement tend vers 0 et donc quand le nombre de symboles envoyés tend vers la dimension m du signal hôte. Le tatouage par étalement de spectre peut donc potentiellement atteindre la limite théorique. De plus, l'utilisation de codes de rendements conventionnels tels que 1/3 ou 1/6 (que nous utiliserons par la suite) donne déjà d'excellentes performances (85 % de la capacité théorique atteignable avec $r = 1/3$).

Même si on peut atteindre théoriquement la capacité maximale en utilisant l'étalement de spectre (figure 1.2(a)), c'est en pratique impossible. Comme vu au-dessus, il faudrait que $n \to m$. Mais le canal ainsi obtenu ne pourraient être gaussien avec un rapport m/n si faible. La courbe de la figure 1.2(b) montre la différence entre capacité maximale théorique (courbe en pointillé, atteignable pour un signal hôte gaussien i.i.d. ou avec des canaux parallèles gaussiens) et capacité en se limitant à $m/n = 25$ (courbe WSS). On voit que les deux mesures se rejoignent pour les attaques les plus fortes, mais la différence est importante pour les attaques faibles. Néanmoins, cela ne concerne que la mesure de capacité et n'a pas d'influence sur la robustesse de la transmission : pour un message de longueur fixée, plus E_b/N_0 (et donc $\mathcal{C}^{max}$) est important et plus la transmission sera robuste.



(a) $n = 200000$ (b) $n = 10000$

FIG. 1.2 – Capacité théorique atteignable par l'étalement de spectre en fonction de la taille du sous-espace ($m \simeq 250000$)

Conclusion

Nous avons considéré une insertion de symboles par étalement de spectre, puis un canal d'attaque de type SAWGN. En posant comme hypothèse un signal hôte modélisé par un ensemble de lois Normales indépendantes (signal gaussien non identiquement distribué), nous avons défini l'estimateur optimal au sens du critère MAP. On peut voir que la transmission de bit *via* ce type de signal hôte et ce type d'attaque correspond à une transmission sur un canal gaussien, dont le rapport signal-à-bruit a été exprimé. Ce rapport permet de quantifier la performance du schéma de tatouage en fonction de la répartition de l'énergie de la marque $\sigma^2_{W_i}$ et des paramètres de l'attaque γ_i et σ_{Z_i}.

Chapitre 2

Résolution par max-min

Nous avons vu dans le chapitre précédent la définition formelle du canal de tatouage en fonction des paramètres d'insertion et d'attaque et nous en avons déduit une mesure de performance, correspondant au rapport signal-à-bruit E_b/N_0.

Afin de trouver la meilleure répartition d'énergie de la marque, nous allons définir la forme de l'attaque optimale en fonction de l'énergie d'insertion. Puis nous verrons la répartition qui permet de maximiser la performance du schéma en prenant en compte l'attaque définie précédemment. Ces recherches se font sous la contrainte de distorsions maximales : l'attaquant ne peut modifier trop fortement les données marquées car le document doit rester utilisable. De même, la marque doit être la moins perceptible possible.

2.1 Le jeu du tatouage

La chaîne du tatouage peut être vue comme un jeu entre deux adversaires : le défenseur qui souhaite transmettre correctement son message, et l'attaquant qui veut au contraire rendre la transmission impossible. La théorie des jeux [Owe95, Sta99] permet de définir des stratégies optimales pour les deux parties. Nous verrons d'abord son principe et quelques exemples simples, puis nous reformulerons le problème du tatouage sous forme de jeu.

2.1.1 Principes

La théorie des jeux est l'étude des comportements rationnels des individus en situation de conflit. Les applications pratiques sont l'économie (déterminer le prix d'un produit en fonction de sa clientèle), les jeux de stratégie (échecs, poker, ...) ou la politique (un exemple souvent cité est celui de Kroutchev et Kennedy lors de la crise de Cuba). Un jeu est un ensemble de règles définissant les gains et les pertes de joueurs en fonction de leurs choix. À chaque tour du jeu, chaque joueur doit choisir une action. Une **stratégie** est l'ensemble des actions effectuées par un joueur lors d'un jeu. Le jeu est à **information complète** si chacun des participants connaît ses possibilités d'action, l'ensemble des choix des autres joueurs, les issues possibles, la valeur des gains qui

	Ciseaux	Feuille	Caillou
Ciseaux	$0, 0$	$-1, +1$	$+1, -1$
Feuille	$+1, -1$	$0, 0$	$-1, +1$
Caillou	$-1, +1$	$+1, -1$	$0, 0$

TAB. 2.1 – Forme normale du jeu *ciseaux feuille caillou*

en résultent et enfin les motifs des joueurs : chacun doit pouvoir se mettre à la place d'un de ses adversaires et savoir quoi décider dans la même situation. Si les joueurs jouent en même temps, il y a **information imparfaite** : un joueur ne sait pas ce que joue son adversaire au moment de son choix. Si les coups sont consécutifs, le jeu est à **information parfaite**.

Un exemple simple de jeu à information imparfaite est *ciseaux/feuille/caillou*. Les deux adversaires annoncent leur choix en même temps et les points sont comptés ainsi : il y a match nul si les joueurs A et B donnent le même objet, le caillou bat les ciseaux, la feuille bat le caillou et les ciseaux battent la feuille. Ces règles peuvent être mises sous la forme d'une matrice, comme le montre la table 2.1. Tout jeu à information complète est présentable sous cette forme. Le jeu perd tout intérêt s'il est à information parfaite : le joueur passant en second est certain de gagner.

Un autre cas très connu est le dilemme du prisonnier. Deux suspects sont arrêtés pour avoir commis un délit. La police manque de preuve et tente d'obtenir des aveux. Les scénarii suivants sont possibles :

- aucun des deux suspects n'avoue. Ils prennent simplement 1 an de prison (on pose arbitrairement le gain à 3 points),
- les deux avouent : ils ont tous les deux 3 ans de prison (1 point),
- l'un des deux avoue et l'autre nie : celui qui a avoué est libre (5 points) et le complice dénoncé prend 5 ans de prison (aucun point).

Ces règles sont résumées par l'arbre de la figure 2.1 (arbre de Kuhn du jeu). Si le jeu est de type coopératif (on recherche un gain collectif maximal), la meilleure solution est que les deux prisonniers nient : ils s'en sortent avec une peine faible. Mais si l'on considère des intérêts individuels (jeu non coopératif), les deux prisonniers vont avouer et dénoncer leur complice. En effet, quel que soit le choix de A, il est plus intéressant pour B d'avouer. Ce raisonnement s'applique aussi pour A et est valable pour un jeu à coups consécutifs ou à coups simultanés. Les prisonniers vont donc tous les deux se dénoncer et écoper d'une forte peine de prison : la rationalité individuelle fait que l'on se trouve dans la pire situation globale (6 ans de prison en tout).

2.1.2 Application au problème du tatouage

Nous définissons une stratégie d'insertion $\mathtt{e}$ comme une répartition de l'énergie moyenne de la marque, c'est-à-dire une suite de σ_{W_i} : $\mathtt{e} = \{\sigma_{W_1}, \sigma_{W_2}, \ldots, \sigma_{W_m}\}$. L'ensemble de toutes les stratégies d'insertion est noté $\mathcal{E} = [\mathbb{R}^+]^m$ et l'ensemble des

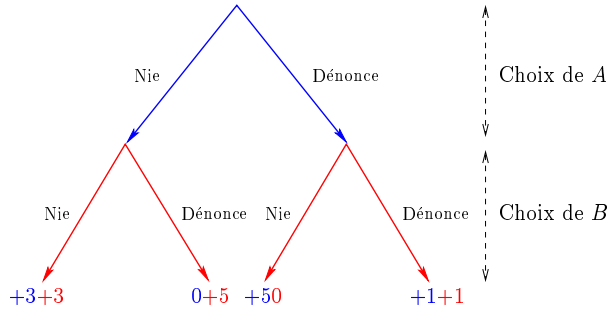


FIG. 2.1 – Arbre de Kuhn du dilemme du prisonnier

stratégies respectant la contrainte de distorsion D_{xy} est noté $\mathcal{E}(D_{\mathsf{xy}})$. De la même façon, une stratégie d'attaque a est une suite de couples (γ_i, σ_{Z_i}). L'ensemble des stratégies d'attaque est $\mathcal{A}$, égal à $(\mathbb{R} \times \mathbb{R}^+)^m$, et celles qui respectent la distorsion maximale D_{a} sont dans $\mathcal{A}(D_{\mathrm{a}})$.

Notre problème est de savoir comment déterminer la stratégie d'insertion. L'interaction entre défenseur (phase d'insertion) et attaquant (phase d'attaque) est un jeu à information complète comprenant un tour unique et où l'attaquant joue après le défenseur (coups consécutifs). Comme nous souhaitons que la technique développée ici respecte le principe de Kerckoffs [Ker83], l'attaquant connaît parfaitement la stratégie d'insertion : le jeu est à information parfaite. Il est en position de force car il peut agir de façon optimale pour contrer la défense. Par contre, le défenseur ne peut prévoir comment vont être attaquées les données marquées. La théorie des jeux nous indique tout de même que l'on peut s'assurer d'une performance minimale, en considérant le pire cas[1]. Du point de vue de la défense, ce pire cas est l'attaque optimale (attaquant omniscient). Si l'on considère que l'attaquant va toujours jouer de la meilleure façon possible, la défense peut définir une stratégie prenant en compte cette attaque. Du point de vue de l'attaquant, résoudre le jeu consiste donc à minimiser la mesure de performance E_b/N_0 (fonction de $\mathcal{A} \times \mathcal{E}$, définie au chapitre précédant) pour une stratégie e donnée, en respectant sa contrainte de distorsion maximale. Sa solution est donc définie par

$$\mathsf{a}_{\mathsf{e}}(D_{\mathrm{a}}) = \arg \min_{\mathsf{a} \in \mathcal{A}(D_{\mathrm{a}})} \left\{ \frac{E_b}{N_0}(\mathsf{e}, \mathsf{a}) \right\}. \tag{2.1}$$

Et pour le défenseur, la stratégie assurant la meilleure performance face à ce type d'attaques est

$$\mathsf{e}(D_{\mathsf{xy}}, D_{\mathrm{a}}) \quad = \quad \arg \max_{\mathsf{e} \in \mathcal{E}(D_{\mathsf{xy}})} \left\{ \frac{E_b}{N_0}(\mathsf{e}, \mathsf{a}_{\mathsf{e}}(D_{\mathrm{a}})) \right\} \tag{2.2}$$

[1] On retrouve ce même type de raisonnement dans les démonstrations de capacité de canal. On considère que la pire dégradation sera appliquée, et la recherche de la meilleure stratégie de défense donne la capacité du canal.

$$= \arg \max_{\mathsf{e} \in \mathcal{E}(D_{\mathsf{xy}})} \left\{ \min_{\mathsf{a} \in \mathcal{A}(D_{\mathrm{a}})} \left\{ \frac{E_b}{N_0} (\mathsf{e}, \mathsf{a}) \right\} \right\}. \tag{2.3}$$

La résolution du jeu, c'est-à-dire la stratégie à utiliser lors de l'insertion, se compose d'une première phase pour rechercher la forme de l'attaque optimale, puis d'une seconde pour trouver la meilleure stratégie de défense (optimisation max-min). Ces optimisations sont contraintes par des mesures de distorsions maximales pour les deux parties.

2.2 Mesures de distorsion

Afin de mesurer les distorsions introduites par l'insertion et l'attaque, nous utilisons une erreur quadratique moyenne pondérée par un facteur perceptuel (mesure de type wEQM comme vu dans la section 2.3.2 de la première partie). La réalisation de W_i n'étant pas forcément connue lors de la répartition d'énergie, nous utilisons l'espérance de l'erreur quadratique. La distorsion d'insertion est alors donnée par

$$D_{\mathsf{xy}} = \frac{1}{m} \mathbb{E} \left[\sum_{i=1}^{m} \varphi_i^2 \left(X_i - Y_i \right)^2 \right], \tag{2.4}$$

où φ_i mesure l'importance perceptuelle du $i^{\text{ème}}$ échantillon du signal hôte x. En reprenant la formulation de y, donnée par les équations 1.1 et 1.2 de la page 64, on obtient

$$D_{\mathsf{xy}} = \frac{1}{m} \sum_{i=1}^{m} \varphi_i^2 \times \sigma_{W_i}^2. \tag{2.5}$$

Comme discuté par Moulin [Mou01] et Cohen [CL02], il existe deux possibilités pour mesurer la distorsion introduite par l'attaque. La première est de considérer l'attaque comme un processus indépendant et alors de prendre uniquement la distorsion introduite par l'attaquant, c'est-à-dire $D_{\mathrm{a}} = D_{\mathsf{yy}'}$. Ce cas supprime une partie des liens entre tatouage et attaque. Par exemple, si l'insertion utilise une mesure de type JND (voir la section 2.3.3 de la première partie) afin de rester sous le seuil de perception, en considérant uniquement la distorsion à partir de y (et donc en calculant de nouveaux seuils de visibilité depuis y), l'attaquant peut très bien faire en sorte de rester sous $\mathrm{nvf}_{\mathsf{y}}$ mais dépasser $\mathrm{nvf}_{\mathsf{x}}$. Les modifications de l'attaque ne seront pas perceptibles si on considère le document marqué, mais pourront être observées si on prend le document original. La seconde possibilité, avec $D_{\mathrm{a}} = D_{\mathsf{xy}'}$, prend en compte l'impact global de la chaîne du tatouage. Ce point de vue est plus en accord avec la notion d'utilisabilité du document final : ce qui importe est la fidélité par rapport au document d'origine, et non par rapport au document marqué. Nous avons choisi cette orientation pour la suite.

Comme le défini le jeu, l'attaquant devra minimiser la performance de la transmission en restant sous un seuil de distorsion maximal, mesuré par $D_{\mathsf{xy}'}$. Mais le signal hôte lui est inconnu. Par contre, il connaît parfaitement la stratégie d'insertion (c'est-à-dire la fonction $f(\sigma_{X_i}) = \sigma_{W_i}$). De plus, comme $\sigma_{Y_i}^2 = \sigma_{X_i}^2 + \sigma_{W_i}^2$ et $\sigma_{W_i}^2 \ll \sigma_{X_i}^2$ du fait des

contraintes d'imperceptibilité, il peut estimer finement les valeurs σ_{X_i} connaissant y (par exemple grâce à un simple algorithme itératif initialisé par $\sigma_{X_i} = \sigma_{W_i}$). Un raisonnement similaire peut s'appliquer sur les pondérations φ_i. En considérant les données à la disposition de l'attaquant, la distorsion s'exprime par

$$\begin{aligned} D_{\mathsf{xy}'} &= \frac{1}{m}\mathbb{E}\left[\sum_{i=1}^{m}\varphi_i^2\left(X_i - Y_i'\right)^2\right] \\ &= \frac{1}{m}\sum_{i=1}^{m}\varphi_i^2\left[(1-\gamma_i)^2\,\mathbb{E}\left[X_i^2\right] + \gamma_i^2\mathbb{E}\left[W_i^2\right] + \mathbb{E}\left[Z_i^2\right]\right] \\ &= \frac{1}{m}\sum_{i=1}^{m}\varphi_i^2\left[\sigma_{X_i}^2(1-\gamma_i)^2 + \gamma_i^2\sigma_{W_i}^2 + \sigma_{Z_i}^2\right]. \end{aligned} \tag{2.6}$$

2.3 Stratégie d'attaque

Connaissant l'expression des attaques possibles et la formule de distorsion, nous allons définir la stratégie d'attaque minimisant la performance. Soit e une répartition de l'énergie de la marque w. La stratégie d'attaque optimale répondant à e et respectant la contrainte $D_{\mathsf{xy}'}^{\max}$ est formulée par

$$\mathsf{a}_{\mathsf{e}}(D_{\mathsf{xy}'}^{\max}) = \arg\min_{\mathsf{a}\in\mathcal{A}(D_{\mathsf{xy}'}^{\max})}\left\{\frac{E_b}{N_0}(\mathsf{e},\mathsf{a})\right\} \tag{2.7}$$

comme vu précédemment. Dans la suite de cette section, nous allons voir comment formuler le problème sous la forme d'une minimisation de fonctionnelle, puis nous verrons sa résolution.

2.3.1 Formulation lagrangienne

L'optimisation de l'équation précédente peut se reformuler sous la forme d'une minimisation lagrangienne. La fonctionnelle à minimiser est alors donnée par la formulation lagrangienne :

$$\begin{aligned} \mathsf{a}_{\mathsf{e}}(D_{\mathsf{xy}'}^{\max}) &= \arg\min_{\mathsf{a}\in\mathcal{A}}\left\{\frac{E_b}{N_0}(\mathsf{a},\mathsf{e}) + \lambda'\left[D_{\mathsf{xy}'} - D_{\mathsf{xy}'}^{\max}\right]\right\} \\ &= \arg\min_{\mathsf{a}\in\mathcal{A}}\left\{J_\lambda = n\frac{E_b}{N_0}(\mathsf{a},\mathsf{e}) + \lambda m\left[D_{\mathsf{xy}'} - D_{\mathsf{xy}'}^{\max}\right]\right\}, \end{aligned} \tag{2.8}$$

où λ est le multiplicateur lagrangien. Les facteurs n et m permettent de simplifier l'expression sans modifier la résolution. Depuis les équations 1.19 et 2.6, on remarque que J_λ est une fonctionnelle additive de la forme

$$J_\lambda = \sum_{i=1}^{m} J_\lambda^i \tag{2.9}$$

$$\text{avec } J_\lambda^i = \frac{\gamma_i^2 \sigma_{W_i}^2}{\gamma_i^2 \left(\sigma_{X_i}^2 + \sigma_{W_i}^2 (n-1)/n\right) + \sigma_{Z_i}^2} + \lambda \varphi_i^2 \left[\sigma_{X_i}^2 \left(1-\gamma_i^2\right) + \gamma_i^2 \sigma_{W_i}^2 + \sigma_{Z_i}^2\right], \tag{2.10}$$

et que donc la minimisation peut se faire terme à terme. La forme de l'attaque optimale au $i^{\text{ème}}$ échantillon est donc donnée par

$$\left(\gamma_i^\star, \sigma_{Z_i}^\star\right) = \arg \min_{\gamma_i, \sigma_{Z_i} \geq 0} \left\{J_\lambda^i\right\}. \tag{2.11}$$

Afin de résoudre cette minimisation, nous recherchons d'abord une solution sur $\mathbb{R} \times \mathbb{R}^{+*}$, puis nous analyserons le comportement de la fonctionnelle aux bords du domaine de validité (ici $\sigma_{Z_i} = 0$). Le cas général se résoud en annulant les dérivées selon chacun des deux paramètres :

$$\frac{\partial J_\lambda^i}{\partial \gamma_i} = 2 \frac{\gamma_i \sigma_{W_i}^2 \sigma_{Z_i}^2}{\left(\gamma_i^2 \left(\sigma_{X_i}^2 + \sigma_{W_i}^2 (n-1)/n\right) + \sigma_{Z_i}^2\right)^2} + 2\lambda \varphi_i^2 \left[\gamma_i \sigma_{W_i}^2 - \sigma_{X_i}^2 (1-\gamma_i)\right] \tag{2.12}$$

$$\frac{\partial J_\lambda^i}{\partial \sigma_{Z_i}^2} = -\frac{\gamma_i^2 \sigma_{X_i}^2}{\left(\gamma_i^2 \left(\sigma_{X_i}^2 + \sigma_{W_i}^2 (n-1)/n\right) + \sigma_{Z_i}^2\right)^2} + \lambda \varphi_i^2 \tag{2.13}$$

La mise à zéro des deux équations donnent les paramètres solutions candidats suivants :

$$\gamma_i^{\mathrm{a}} = n \frac{\sqrt{\lambda} \varphi_i \sigma_{X_i}^2 - \sigma_{W_i}}{\sqrt{\lambda} \varphi_i \sigma_{W_i}^2} \tag{2.14}$$

$$\sigma_{Z_i}^{\mathrm{a}} = \sqrt{\gamma_i^{\mathrm{a}} \left(\gamma_i^{\mathrm{w}} - \gamma_i^{\mathrm{a}}\right) \left(\sigma_{X_i}^2 + \sigma_{W_i}^2\right)}, \tag{2.15}$$

où γ_i^{w} est l'expression du facteur multiplicatif utilisé pour un filtrage de Wiener dont le but serait de débruiter le signal hôte du signal de la marque. Il est défini par

$$\gamma_i^{\mathrm{w}} = \frac{\sigma_{X_i}^2}{\sigma_{X_i}^2 + \sigma_{W_i}^2}. \tag{2.16}$$

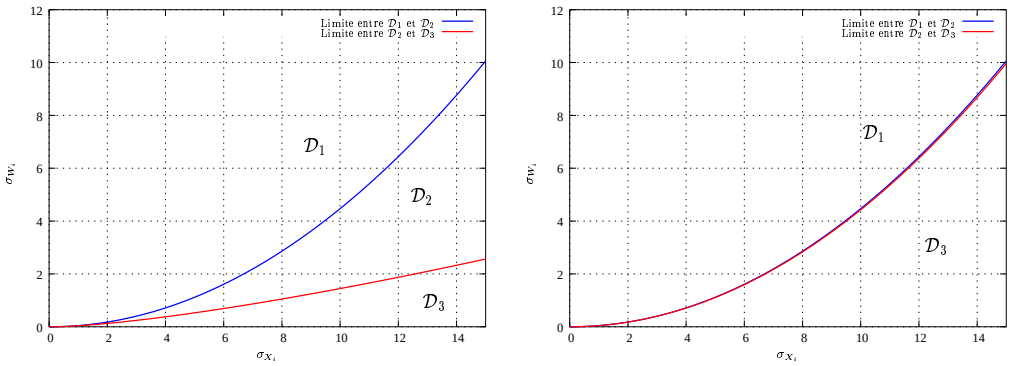
Cette stratégie d'attaque est un mélange de filtrage et d'ajout de bruit gaussien. Nous la notons $\mathsf{a_l}(i) = (\gamma_i^{\mathrm{a}}, \sigma_{Z_i}^{\mathrm{a}})$.

2.3.2 Cas limites

Par définition, la valeur de σ_{Z_i} est contrainte : elle doit être supérieure ou égale à zéro. Les solutions des équations 2.14 et 2.15 ne sont valides que si les inégalités suivantes sont vérifiées :

$$\sqrt{\lambda} \varphi_i \sigma_{X_i}^2 - \sigma_{W_i} \geq 0 \tag{2.17}$$

$$\left(\sigma_{W_i} - \sqrt{\lambda} \varphi_i \sigma_{X_i}^2\right) \left(\sigma_{X_i}^2 + \sigma_{W_i}^2\right) + \frac{\sqrt{\lambda} \varphi_i \sigma_{X_i}^2 \sigma_{W_i}^2}{n} \geq 0 \tag{2.18}$$



(a) Avec $n = 1$

(b) Avec $n = 100$: le domaine $\mathcal{D}_2$ est quasi-nul

FIG. 2.2 – Les trois domaines d'attaque définis par la contrainte $\sigma_{Z_i} \geq 0$ ($\lambda = 0,02$ et $\varphi_i = 1$)

Ces deux inégalités définissent trois domaines, représentés par la figure 2.2, et notés $\mathcal{D}_1$ (contrainte 2.17 non respectée), $\mathcal{D}_2$ (domaine de validité des solutions trouvées précédemment) et $\mathcal{D}_3$ (contrainte 2.18 non satisfaite). Les domaines $\mathcal{D}_1$ et $\mathcal{D}_3$ admettent une attaque optimale autre que celle définie par l'annulation des dérivées. Nous considérons alors la limite de validité $\sigma_{Z_i} = 0$ et recherchons les valeurs de γ_i susceptibles de donner un minimum local. D'après la dérivée de la fonctionnelle par rapport à γ_i (équation 2.12), deux valeurs sont alors possibles :

- $\gamma_i = \gamma_i^{\mathrm{W}}$ (équation 2.16). Cette stratégie est obtenue par annulation de la dérivée. Elle est notée $\mathsf{a}_{\mathsf{W}}(i) = (\gamma_i^{\mathrm{W}}, 0)$,
- $\gamma_i = 0$, la valeur marquée est annulée. C'est un cas limite. Cette stratégie est notée $\mathsf{a}_{\mathsf{E}}(i) = (0, 0)$.

Les trois attaques possibles aboutissent aux valeurs de fonctionnelles J_λ^i données par la table 2.2. Afin de trouver le candidat optimal, il faut rechercher quelle stratégie parmi les trois possibles minimise J_λ^i dans chacun des trois domaines. Grâce aux valeurs des fonctionnelles données par la table 2.2 et avec la définition des domaines (équations 2.17 et 2.18), on montre que les stratégies optimales pour les domaines $\mathcal{D}_1$, $\mathcal{D}_2$ et $\mathcal{D}_3$ sont respectivement a_{E}, a_{I} et a_{W}. Le détail des calculs est donné dans l'annexe A.2.1.

2.3.3 Remarques

On peut analyser intuitivement le comportement de l'attaquant. Ainsi, si l'énergie de la marque est importante par rapport à celle du signal hôte (domaine $\mathcal{D}_1$), il est plus intéressant d'annuler le signal : même si la distorsion introduite est forte, la perte en terme de E_b/N_0 sera très importante. À l'opposé, si le signal hôte est fort et que la marque est faible (domaine $\mathcal{D}_3$), annuler le signal (et supprimer la marque) apporterait beaucoup de distorsion pour une perte de performance trop faible : il vaut mieux alors

$\mathsf{a}_{\mathsf{E}}(i) = (0,0)$	J_E	$=$	$\lambda\varphi_i^2\sigma_{X_i}^2$
$\mathsf{a}_{\mathsf{W}}(i) = (\gamma_i^{\mathrm{w}}, 0)$	J_W	$=$	$\sqrt{\lambda}\varphi_i \dfrac{\sigma_{X_i}^2\sigma_{W_i}}{\sigma_{X_i}^2+\sigma_{W_i}^2} + \lambda\varphi_i^2 \dfrac{\sigma_{X_i}^2\sigma_{W_i}^2}{\sigma_{X_i}^2+\sigma_{W_i}^2}$
$\mathsf{a}_{\mathsf{I}}(i) = (\gamma_i^{\mathrm{a}}, \sigma_{Z_i}^{\mathrm{a}})$	J_I	$=$	$\gamma_i^{\mathrm{a}}\sqrt{\lambda}\varphi_i\sigma_{W_i} + \lambda\varphi_i^2\sigma_{X_i}^2\left(1-\gamma_i^{\mathrm{a}}\right)$

TAB. 2.2 – Valeur de la fonctionnelle J_λ^i à minimiser en fonctions des trois attaques possibles

laisser la marque et simplement utiliser un filtre de Wiener pour abaisser $D_{\mathsf{xy}'}$.

Enfin, en posant

$$\begin{aligned}\frac{\sigma_{X_i}^2}{\sigma_{X_i}^2+\sigma_{W_i}^2+\left(\sigma_{Z_i}^{\mathrm{a}}/\gamma_i^{\mathrm{a}}\right)^2} &= \frac{\sigma_{X_i}^2}{\sigma_{X_i}^2+\sigma_{W_i}^2+\left(\sigma_{X_i}^2+\sigma_{W_i}^2\right)\left(\gamma_i^{\mathrm{w}}-\gamma_i^{\mathrm{a}}\right)/\gamma_i^{\mathrm{a}}} \\ &= \gamma_i^{\mathrm{a}}, \end{aligned} \tag{2.19}$$

on voit que γ_i^{a} est l'expression du facteur multiplicatif d'un filtre de Wiener dont le but serait d'atténuer la marque et un bruit d'attaque d'énergie $\left(\sigma_{Z_i}^{\mathrm{a}}/\gamma_i^{\mathrm{a}}\right)^2$. L'attaque intermédiaire consiste donc à ajouter du bruit et à débruiter la marque et le bruit afin de restaurer le signal hôte original. Ces résultats sont à rapprocher de ceux de Cohen et Lapidoth obtenus par une optimisation max-min similaire [CL02] (expliqués dans la section 4.3 de la première partie) et à ceux de Su *et al.*, résumés par : *at low distortions, add noise ; at high distortions, throw away frequency components* [SEG01b].

2.4 Défense

Comme nous considérons que l'attaquant est susceptible d'appliquer la stratégie optimale vue dans la section précédente, afin d'assurer d'une performance minimale, nous devons définir la stratégie d'insertion maximisant la performance de la transmission, c'est-à-dire

$$\mathsf{e}(D_{\mathsf{xy}}^{\max}, D_{\mathsf{xy}'}^{\max}) = \arg\max_{\mathsf{e}\in\mathcal{E}(D_{\mathsf{xy}}^{\max})} \left\{ \frac{E_b}{N_0}\left(\mathsf{e}, \mathsf{a}_{\mathsf{e}}(D_{\mathsf{xy}'})\right)\right\}. \tag{2.20}$$

Comme pour la résolution de l'attaque, on utilise une formulation lagrangienne afin de se ramener à la maximisation d'une fonctionnelle :

$$\mathsf{e}(D_{\mathsf{xy}}^{\max}, D_{\mathsf{xy}'}^{\max}) = \arg\max_{\mathsf{e}\in\mathcal{E}} \left\{ J_\chi = J_\lambda - \chi m\left[D_{\mathsf{xy}} - D_{\mathsf{xy}}^{\max}\right]\right\}, \tag{2.21}$$

où χ est le multiplicateur lagrangien. Là encore, on peut remarquer que J_χ est une fonctionnelle additive, définie par la somme des J_χ^i. La valeur optimale de l'énergie de la marque pour le $i^{\text{ème}}$ échantillon est donnée par

$$\sigma_{W_i}^\star = \arg\max_{\sigma_{W_i}\geq 0} \left\{ J_\chi^i \right\}. \tag{2.22}$$

2.4.1 Réponse à l'annulation

Nous considérons les trois attaques possibles et les trois expressions de J^i_λ correspondantes. Dans le cas de l'attaque par effacement $\mathsf{a_E}(i)$ avec $\gamma_i = \sigma_{Z_i} = 0$, optimale sur $\mathcal{D}_1$, nous avons d'après la table 2.2 et l'équation 2.5 :

$$J^i_\chi = \lambda\varphi_i^2\sigma^2_{X_i} - \chi\varphi_i^2\sigma^2_{W_i}. \tag{2.23}$$

Cette fonction est décroissante, et la valeur maximale de la fonctionnelle est donc la borne minimale du domaine $\mathcal{D}_1$, c'est-à-dire $\sigma_{W_i} = \sqrt{\lambda}\varphi_i\sigma^2_{X_i}$. Nous notons cette première stratégie d'insertion $\mathsf{e_E}(i) = \sqrt{\lambda}\varphi_i\sigma^2_{X_i}$. Sa forme est donnée par la courbe bleue de la figure 2.3(a). Les points solutions sont sur la frontière délimitant $\mathcal{D}_1$ et $\mathcal{D}_2$.

2.4.2 Réponse à l'attaque intermédiaire

Considérons la valeur $\sigma_{W_i} = \sqrt{\lambda}\varphi_i\sigma^2_{X_i}(n-1)/n$. Cette valeur ne vérifie pas la contrainte de l'équation 2.18, mais elle vérifie celle de l'équation 2.17. On peut donc dire que $\mathcal{D}_2$ est inclus dans le domaine défini par

$$\left\{\sigma_{W_i} \leq \sqrt{\lambda}\varphi_i\sigma^2_{X_i}\right\} \cap \left\{\sigma_{W_i} \geq \sqrt{\lambda}\varphi_i\sigma^2_{X_i}(n-1)/n\right\}.$$

La solution optimale face à l'attaque intermédiaire appartient à ce domaine. Or il est en pratique extrêmement réduit : les tailles de message utilisées sont souvent supérieures à 50, et dans ce cas $(n-1)/n$ est très proche de 1. Le domaine $\mathcal{D}_2$ est par conséquent encore plus petit. Cela est vérifiable sur la figure 2.2(b). La solution de défense dans ce domaine peut donc s'approximer par

$$\mathsf{e_I}(i) \simeq \sqrt{\lambda}\varphi_i\sigma^2_{X_i}. \tag{2.24}$$

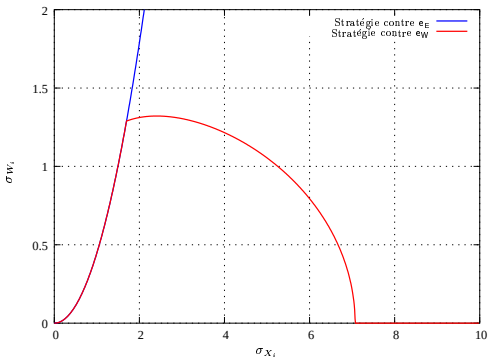
On retrouve la même défense que dans le cas de l'attaque par annulation.

2.4.3 Réponse au filtrage de Wiener

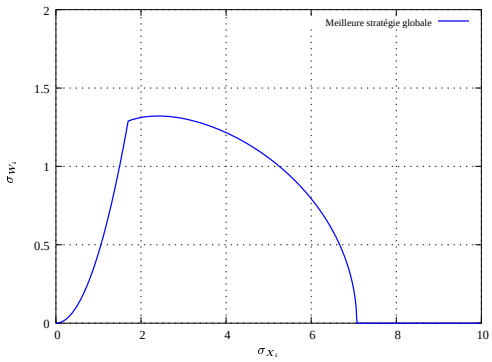
Enfin, pour l'attaque par filtrage de Wiener $\mathsf{a_W}(i) = (\gamma_i^{\mathrm{W}}, 0)$, optimale dans le domaine $\mathcal{D}_3$, nous procédons de la même façon : nous reprenons la valeur de J^i_λ de la table 2.2 et nous recherchons un maximum local par annulation de la dérivée. Néanmoins, cette technique aboutit à la recherche de la racine d'une équation du cinquième degré. Afin de fournir une solution analytique, nous faisons l'hypothèse que $(n-1)/n \simeq 1$ pour les valeurs de n généralement utilisées dans le cadre du tatouage. On a alors la fonctionnelle suivante :

$$J^i_\chi \simeq \widetilde{J}^i_\chi = \frac{\sigma^2_{W_i}}{\sigma^2_{X_i}+\sigma^2_{W_i}}\left[1+\lambda\varphi_i^2\sigma^2_{X_i}\right] - \chi\varphi_i^2\sigma^2_{W_i} \tag{2.25}$$

$$\text{et } \frac{\partial\widetilde{J}^i_\chi}{\partial\sigma^2_{W_i}} = \frac{\sigma^2_{X_i}}{\left(\sigma^2_{X_i}+\sigma^2_{W_i}\right)^2}\left[1+\lambda\varphi_i^2\sigma^2_{X_i}\right] - \chi\varphi_i^2. \tag{2.26}$$



(a) Stratégies d'insertion contre l'attaque par effacement et par filtre de Wiener



(b) Stratégie globale, obtenue en sélectionnant la meilleure valeur de J_χ^i

FIG. 2.3 – Stratégie d'insertion optimale ($\lambda = 0,02$, $\chi = 0,022$, $\varphi_i = 1$ et $n = 100$)

La racine de la dérivée est alors

$$\sigma_{W_i}^2 = \frac{\sigma_{X_i}\sqrt{1+\lambda\varphi_i^2\sigma_{X_i}^2}}{\sqrt{\chi}\varphi_i} - \sigma_{X_i}^2. \tag{2.27}$$

Comme $\sigma_{W_i}^2$ doit être positif ou nul, l'expression ci-dessus n'est valide que si $\varphi_i^2\sigma_{X_i}^2(\lambda-\chi)+1 \geq 0$. Dans le cas contraire, la dérivée est négative et le maximum est obtenu avec $\sigma_{W_i}^2 = 0$. Cette stratégie d'insertion est notée $\mathsf{e_W}(i)$ et est illustrée par la courbe rouge de la figure 2.3(a). De plus, si cette racine donne une valeur supérieure à $\sqrt{\lambda}\varphi_i\sigma_{X_i}^2$ (limite inférieure approximée de $\mathcal{D}_2$), nous ne sommes plus dans le domaine $\mathcal{D}_3$ pour lequel cette solution a été calculée. La stratégie adaptée à l'attaque par filtrage de Wiener est donc définie par

$$\begin{aligned} \mathsf{e_W}(i)^2 &= \min\left[\frac{\sigma_{X_i}\sqrt{1+\lambda\varphi_i^2\sigma_{X_i}^2}}{\sqrt{\chi}\varphi_i} - \sigma_{X_i}^2,\ \mathsf{e_I}(i)^2\right] \text{ si } \varphi_i^2\sigma_{X_i}^2(\lambda-\chi) \geq -1 \\ &= 0 \text{ sinon.} \end{aligned} \tag{2.28}$$

La stratégie d'insertion globale consiste à faire le choix entre les stratégies $\mathsf{e_E}$ (équivalente à $\mathsf{e_I}$) et $\mathsf{e_W}$. Pour cela, il suffit de calculer les valeurs de J_χ^i correspondantes et de sélectionner le meilleur résultat, et ce pour chacun des m éléments du signal hôte x. La figure 2.3(b) montre la forme de la stratégie globale. On peut remarquer que pour les éléments du signal hôte de forte énergie, il vaut mieux ne pas insérer de marque.

Conclusion

En prenant l'hypothèse d'un signal hôte non i.i.d. gaussien, d'un principe de tatouage par étalement de spectre et d'attaques de type SAWGN, nous avons montré une répartition de l'énergie de la marque permettant d'assurer une performance minimale de transmission (rapport signal-à-bruit) quelle que soit l'attaque subie. Pour cela, nous avons exhibé la forme d'attaque la plus néfaste, mélange de trois stratégies (annulation du signal, débruitage par Wiener et combinaison de bruit et de filtrage), en reformulant le problème sur la forme d'une minimisation lagrangienne. Puis nous avons vu comment répartir l'énergie de la marque en prenant en compte une attaque potentiellement optimale.

Chapitre 3

Tatouage additif filtré

La résolution du jeu présentée dans le chapitre précédent nous montre la forme de l'attaque optimale. Une de ses composantes est le filtrage de Wiener : l'attaquant restaure l'image marquée en atténuant le bruit que représente la marque. Il va ainsi réduire la distorsion entre l'image originale et l'image marquée puis attaquée. On peut donc obtenir des cas de transmission tels que $D_{\mathsf{xy}'} < D_{\mathsf{xy}}$: l'image marquée puis attaquée est plus fidèle que l'image non attaquée. De plus, si l'on reprend l'équation du rapport signal-à-bruit

$$\frac{E_b}{N_0} = \frac{1}{m}\sum_{i=1}^{m} \frac{\gamma_i^2 \sigma_{W_i}^2}{\gamma_i^2 \left(\sigma_{X_i}^2 + \sigma_{W_i}^2 (n-1)/n\right) + \sigma_{Z_i}^2}, \tag{3.1}$$

comme les formules de l'attaque optimale nous indiquent que si $\gamma_i = \gamma_i^{\mathrm{W}}$, alors $\sigma_{Z_i} = 0$, la performance de la transmission reste constante. Le filtrage de Wiener réduit la distorsion apportée par l'insertion de la marque, mais ne modifie en rien les performances du schéma de tatouage. Ce filtrage est donc une amélioration possible de la phase d'insertion, sans aucune contrepartie. C'est pour cette raison qu'il est logique de l'introduire lors de l'insertion. Nous développons donc une technique de tatouage **additive filtrée**. Ce nouveau type d'insertion modifie légèrement les formulations de y et y'. Nous posons $y_i = \bar{\gamma}_i (x_i + w_i)$ et $y'_i = \bar{\bar{\gamma}}_i \times y_i + z_i$.

Nous allons résoudre le jeu en prenant en compte la nouvelle expression de l'insertion. Nous verrons d'abord la nouvelle expression des mesures de distorsion, puis nous exprimerons les stratégies d'attaque et de défense.

3.1 Reprise des résultats précédents

Les hypothèses de départ sont inchangées, c'est à dire que chaque élément x_i du signal hôte x est la réalisation de $X_i \sim \mathcal{N}(0, \sigma_{X_i}^2)$ et que le signal de marque est par construction modélisé par $W_i \sim \mathcal{N}(0, \sigma_{W_i}^2)$. Afin de pouvoir reprendre la majorité des résultats des deux précédentes sections, nous notons $\gamma_i = \bar{\gamma}_i \times \bar{\bar{\gamma}}_i$ pour retomber sur l'expression de l'équation 1.4 (page 65), c'est-à-dire $y'_i = \gamma_i (x_i + w_i) + z_i$. De ce fait,

l'expression de l'estimateur optimal selon le MAP est donc toujours

$$\widehat{b_j} = \sum_{i=1}^{m} \frac{\gamma_i \sigma_{W_i} \times \mathsf{G}(i,j) \times y'_i}{\gamma_i^2 \left(\sigma_{X_i}^2 + \sigma_{W_i}^2 (n-1)/n\right) + \sigma_{Z_i}^2}, \tag{3.2}$$

donnant le rapport signal-à-bruit de l'équation 3.1. La distorsion d'attaque $D_{\mathsf{xy}'}$ reste inchangée. Seule la distorsion d'insertion varie par rapport aux mesures de la section 2.2. L'erreur quadratique moyenne pondérée est alors donnée par

$$\begin{aligned} D_{\mathsf{xy}} &= \frac{1}{m}\mathbb{E}\left[\sum_{i=1}^{m} \varphi_i^2 (X_i - Y_i)^2\right] \\ &= \frac{1}{m}\sum_{i=1}^{m} \varphi_i^2 \left[(1-\bar{\gamma}_i)^2 \mathbb{E}\left[X_i^2\right] + \bar{\gamma}_i^2 \mathbb{E}\left[W_i^2\right]\right] \\ &= \frac{1}{m}\sum_{i=1}^{m} \varphi_i^2 \left[\sigma_{X_i}^2 (1-\bar{\gamma}_i)^2 + \bar{\gamma}_i^2 \sigma_{W_i}^2\right]. \end{aligned} \tag{3.3}$$

3.2 Résolution du jeu

Une stratégie d'insertion e est une répartition de l'énergie de la marque sur les échantillons du signal hôte, associée à une suite de pondérations : $\mathsf{e} \in \mathcal{E} = (\mathbb{R} \times \mathbb{R}^+)^m$. L'ensemble des stratégies d'insertion respectant la contrainte de distorsion maximale D_{xy} est noté $\mathcal{E}(D_{\mathsf{xy}})$. Une stratégie d'attaque est une suite de couples $(\bar{\bar{\gamma}}_i, \sigma_{Z_i})$. L'ensemble des stratégies d'attaque est $\mathcal{A}$, et celles qui respectent la distorsion d'attaque maximale $D_{\mathsf{xy}'}$ définissent l'ensemble $\mathcal{A}(D_{\mathsf{xy}'})$.

3.2.1 Stratégie d'attaque

Comme pour la présentation du jeu de la page 71, on suppose que l'attaquant connaît parfaitement la stratégie d'insertion adoptée. Nous allons donc définir la stratégie d'attaque optimale en fonction d'une stratégie e quelconque, respectant la contrainte de distorsion maximale imposée, ce qui revient à résoudre la minimisation lagrangienne

$$\begin{aligned} \mathsf{a}_{\mathsf{e}}(D_{\mathsf{xy}'}^{\max}) &= \arg\min_{\mathsf{a}\in\mathcal{A}} \left\{ \frac{E_b}{N_0}(\mathsf{a},\mathsf{e}) + \lambda' \left[D_{\mathsf{xy}'} - D_{\mathsf{xy}'}^{\max}\right] \right\} \\ &= \arg\min_{\mathsf{a}\in\mathcal{A}} \left\{ J_\lambda = n\frac{E_b}{N_0}(\mathsf{a},\mathsf{e}) + \lambda m \left[D_{\mathsf{xy}'} - D_{\mathsf{xy}'}^{\max}\right] \right\}. \end{aligned} \tag{3.4}$$

En reprenant les étapes des sections 2.3.1 et 2.3.2, on sépare l'espace des paramètres en trois domaines, définis par

$$\begin{aligned} \mathcal{D}_1 &= \left\{ \sqrt{\lambda}\varphi_i \sigma_{X_i}^2 - \sigma_{W_i} < 0 \right\} \\ \mathcal{D}_2 &= \left\{ \sqrt{\lambda}\varphi_i \sigma_{X_i}^2 - \sigma_{W_i} \geq 0 \right\} \end{aligned} \tag{3.5}$$

$$\cap \left\{ \left(\sigma_{W_i} - \sqrt{\lambda}\varphi_i \sigma_{X_i}^2\right)\left(\sigma_{X_i}^2 + \sigma_{W_i}^2\right) + \frac{\sqrt{\lambda}\varphi_i \sigma_{X_i}^2 \sigma_{W_i}^2}{n} \geq 0 \right\} \tag{3.6}$$

$$\mathcal{D}_3 = \left\{ \left(\sigma_{W_i} - \sqrt{\lambda}\varphi_i \sigma_{X_i}^2\right)\left(\sigma_{X_i}^2 + \sigma_{W_i}^2\right) + \frac{\sqrt{\lambda}\varphi_i \sigma_{X_i}^2 \sigma_{W_i}^2}{n} < 0 \right\}. \tag{3.7}$$

On retrouve alors les trois stratégies : annulation ($\mathsf{a}_\mathsf{E}(i) = (\bar{\bar{\gamma}}_i = 0, \sigma_{Z_i} = 0)$), filtrage de Wiener ($\mathsf{a}_\mathsf{W}(i) = (\bar{\bar{\gamma}}_i = \gamma_i^{\mathrm{W}}/\bar{\gamma}_i), \sigma_{Z_i} = 0)$) et mélange de filtrage et bruit ($\mathsf{a}_\mathsf{I}(i) = (\bar{\bar{\gamma}}_i^{\mathrm{a}}, \sigma_{Z_i}^{\mathrm{a}})$), dont les paramètres sont donnés par

$$\bar{\bar{\gamma}}_i^{\mathrm{a}} = n\frac{\sqrt{\lambda}\varphi_i \sigma_{X_i}^2 - \sigma_{W_i}}{\bar{\gamma}_i \sqrt{\lambda}\varphi_i \sigma_{W_i}^2} \tag{3.8}$$

$$\sigma_{Z_i}^{\mathrm{a}} = \sqrt{\bar{\gamma}_i \bar{\bar{\gamma}}_i \left(\gamma_i^{\mathrm{W}} - \bar{\gamma}_i \bar{\bar{\gamma}}_i\right)\left(\sigma_{X_i}^2 + \sigma_{W_i}^2\right)}. \tag{3.9}$$

Ces trois stratégies sont respectivement optimales dans les domaines $\mathcal{D}_1$, $\mathcal{D}_3$ et $\mathcal{D}_2$, et aboutissent aux fonctionnelles J_λ^i de la table 2.2. Comme dans le chapitre précédent, on voit que le domaine $\mathcal{D}_2$ est écrasé pour des valeurs de n réalistes (supérieures à 50).

3.2.2 Stratégie de défense

Trouver la stratégie d'insertion permettant de répondre de façon optimale à l'attaque définie dans la section au-dessus consiste à rechercher pour tout i les paramètres $\bar{\gamma}_i$ et σ_{W_i} maximisant la fonctionnelle

$$\left(\bar{\gamma}_i^\star, \sigma_{W_i}^\star\right) = \arg \max_{\bar{\gamma}_i, \sigma_{W_i} \geq 0} \left\{ J_\chi^i = J_\lambda^i - \chi\varphi_i^2 \left[\sigma_{X_i}^2 \left(1 - \bar{\gamma}_i\right)^2 + \bar{\gamma}_i^2 \sigma_{W_i}^2\right] \right\}. \tag{3.10}$$

Défense face à l'annulation

Face à la stratégie d'attaque a_E, la dérivée suivant $\bar{\gamma}_i$ est donnée par

$$\frac{\partial J_\chi^i}{\partial \bar{\gamma}} = -2\chi\varphi_i^2 \left[\sigma_{X_i}^2 \left(\bar{\gamma}_i - 1\right) + \bar{\gamma}_i \sigma_{W_i}^2\right], \tag{3.11}$$

et est annulée par $\bar{\gamma}_i = \gamma_i^{\mathrm{W}}$. De plus, comme $\bar{\gamma}_i$ est indépendant de J_λ^i, ce résultat est valable pour les trois cas (annulation, intermédiaire et filtrage de Wiener). La dérivée selon $\sigma_{W_i}^2$ est négative, et donc la valeur maximisant la fonctionnelle est la limite inférieure du domaine $\mathcal{D}_1$, c'est-à-dire $\sigma_{W_i} = \sqrt{\lambda}\varphi_i \sigma_{X_i}^2$. La stratégie optimale dans ce cas de figure est donc $\mathsf{e}_\mathsf{E}(i) = (\gamma_i^{\mathrm{W}}, \sqrt{\lambda}\varphi_i \sigma_{X_i}^2)$. Elle est illustrée par la courbe bleue de la figure 3.1(a).

Face à l'attaque intermédiaire

Nous avons vu dans le chapitre précédent que le domaine $\mathcal{D}_2$ était en pratique très réduit. Nous avions montré que si il existait une solution autre que les limites du domaine, elle était comprise entre $\sqrt{\lambda}\varphi_i \sigma_{X_i}^2$ et $\sqrt{\lambda}\varphi_i \sigma_{X_i}^2 (n-1)/n$. Si cette solution n'existe pas, la meilleure stratégie consiste à prendre une valeur de σ_{W_i} sur les bords. La solution peut alors être approximée par $\sigma_{W_i} \simeq \sqrt{\lambda}\varphi_i \sigma_{X_i}^2$, c'est-à-dire $\mathsf{e}_\mathsf{I}(i) \simeq \mathsf{e}_\mathsf{E}(i)$.

Défense face au filtrage de Wiener

En reprenant l'astuce de la section 2.4.3 (c'est-à-dire $(n-1)/n \simeq 1$ pour des tailles de message usuelles), on calcule la dérivée selon $\sigma^2_{W_i}$:

$$\frac{\partial J^i_\chi}{\partial \sigma^2_{W_i}} \simeq \frac{\sigma^2_{X_i}}{\left(\sigma^2_{X_i} + \sigma^2_{W_i}\right)^2}\left[1 + \lambda\varphi_i^2\sigma^2_{X_i}\right] - \chi\varphi_i^2\bar{\gamma}_i^2. \tag{3.12}$$

Or, nous avons vu que $\bar{\gamma}_i$ est l'expression d'un filtre de Wiener. Nous pouvons l'introduire dans l'équation au-dessus, et nous obtenons

$$\frac{\partial J^i_\chi}{\partial \sigma^2_{W_i}} \simeq \frac{\sigma^2_{X_i}}{\left(\sigma^2_{X_i} + \sigma^2_{W_i}\right)^2}\left[1 + \varphi_i^2\sigma^2_{X_i}(\lambda - \chi)\right]. \tag{3.13}$$

Si $1 + \varphi_i^2\sigma^2_{X_i}(\lambda - \chi) < 0$, la fonction J^i_χ est décroissante et la valeur optimale de σ_{W_i} est donc 0. Sinon, la fonction est croissante, et la valeur optimale est la limite supérieure du domaine $\mathcal{D}_3$, qui correspond également à la limite inférieure de $\mathcal{D}_2$, que nous avons approximée par $\sigma_{W_i} \simeq \sqrt{\lambda}\varphi_i\sigma^2_{X_i}$. La stratégie optimale face au filtrage de Wiener est donc

$$\begin{aligned} \mathsf{e}_\mathsf{W}(i) &= (\bar{\gamma}_i = \gamma_i^\mathrm{W}, \sigma_{W_i} = \sqrt{\lambda}\varphi_i\sigma^2_{X_i}) \text{ si } \sigma^2_{X_i} \leq \sigma^2_{\lim} \\ &= (\gamma_i^\mathrm{W}, 0) \text{ sinon.} \end{aligned} \tag{3.14}$$

avec $\sigma^2_{\lim} = \left[\varphi_i^2(\chi - \lambda)\right]^{-1}$. Elle est illustrée par la courbe rouge de la figure 3.1(a).

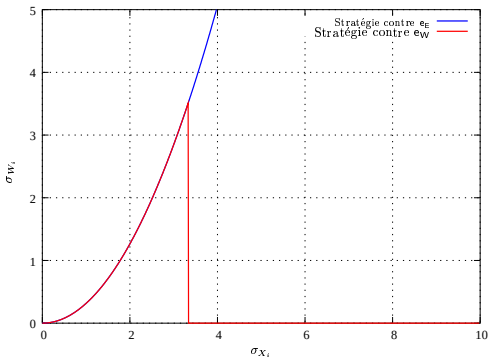
Choix global

Pour savoir quelle stratégie de défense est préférable entre e_E (équivalente à e_I) et e_W, nous reprenons la formulation de J^i_χ, en introduisant le fait que $\bar{\gamma}_i = \gamma_i^\mathrm{W}$:

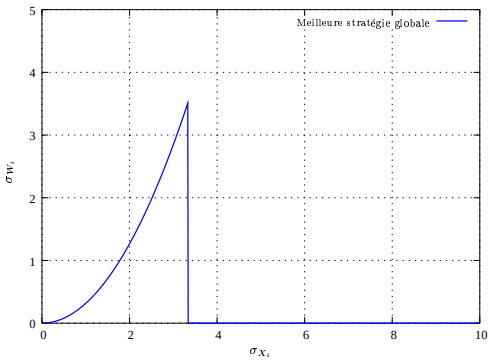
$$J^i_\chi = J^i_\lambda - \chi\varphi_i^2\frac{\sigma^2_{X_i}\sigma^2_{W_i}}{\sigma^2_{X_i} + \sigma^2_{W_i}}. \tag{3.15}$$

Si $\sigma^2_{X_i} \leq \sigma^2_{\lim}$, les fonctionnelles sont égales pour toutes les stratégies (voir les équations A.10 à A.14 de l'annexe A.2.1). Par contre, si $\sigma^2_{X_i} > \sigma^2_{\lim}$, les fonctionnelles sont négatives sauf pour $\sigma_{W_i} = 0$, c'est-à-dire pour la stratégie e_W. C'est donc elle qui maximise J^i_χ. Au final, cette stratégie est la meilleure pour toutes les valeurs de σ_{X_i}. Elle est illustrée par la figure 3.1(b).

On remarque également que l'énergie de la marque est croissante en φ_i : plus l'échantillon est perceptuellement important et plus il sera marqué. L'intuition de Cox [CKLS97] que nous avons évoquée dans la section 4.3 (partie 1) est en partie vérifiée.



(a) Stratégies d'insertion contre l'attaque par effacement et par filtre de Wiener



(b) Stratégie globale, obtenue en sélectionnant la meilleure valeur de J_χ^i

FIG. 3.1 – Stratégie d'insertion optimale ($\lambda = 10^{-4}$, $\chi = 10^{-3}$, $\varphi_i = 1$ et $n = 100$)

3.3 Mise en œuvre pratique

Nous venons de montrer que la répartition optimale de la marque est donnée par $\sigma_{W_i} \simeq \sqrt{\lambda}\varphi_i\sigma_{X_i}^2$ pour tout i respectant $\sigma_{X_i}^2 \leq \sigma_{\text{lim}}^2$, ce qui correspond exactement à la frontière entre les deux stratégies d'attaque possibles (annulation de l'échantillon en posant $\bar{\bar{\gamma}}_i = 0$ ou aucune action). Il nous est donc impossible de connaître le choix de l'attaquant lors de la réception des données marquées y. S'il décide que la $i^{\text{ème}}$ donnée appartient au domaine $\mathcal{D}_1$, il l'annule et dans ce cas

$$J_\lambda^i = J_E = \lambda\varphi_i^2\sigma_{X_i}^2. \tag{3.16}$$

S'il décide qu'elle appartient à $\mathcal{D}_3$, $\bar{\bar{\gamma}}_i = 1$ et $\sigma_{Z_i} = 0$. Nous avons alors $J_\lambda^i = J_W = J_E$ d'après l'équation A.13 de l'annexe A.2.1. Enfin, s'il décide que la donnée marquée est dans le domaine $\mathcal{D}_2$, il utilisera l'attaque intermédiaire. Mais aussi dans ce cas, nous avons $J_\lambda^i = J_I = J_E$ (équation A.11 de l'annexe A.2.1). Quelle que soit la stratégie d'attaque choisie (entre les trois stratégies optimales), la fonctionnelle que l'attaquant cherche à minimiser donne le même coût. On peut donc en déduire que

$$J_E = J_I = J_W = \lambda\varphi_i^2\sigma_{X_i}^2 \quad \Rightarrow \quad \frac{E_b}{N_0} = \lambda\left[-mD_{\mathsf{xy}'} + \sum_{i=1}^{\sigma_{X_i}\leq\sigma_{\text{lim}}} \varphi_i^2\sigma_{X_i}^2\right]. \tag{3.17}$$

À partir d'un couple (λ, χ)[1], on peut donc tracer une ligne indiquant le rapport signal-à-bruit E_b/N_0 minimum qu'il est possible d'atteindre, pour des distorsions d'attaque comprises entre D_{xy} et la distorsion suite à l'annulation de tous les échantillons marqués. Deux exemples sont donnés par la figure 3.2. Les couples (λ, χ) ont été choisis afin que la

[1] Ou un couple $(\lambda, \sigma_{\text{lim}})$.

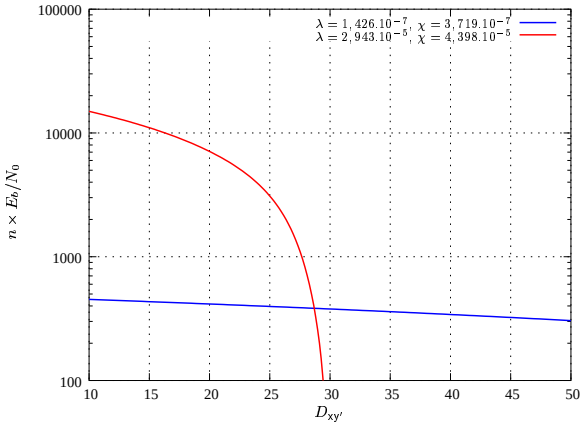


FIG. 3.2 – Performances obtenues avec deux couples (λ, χ) différents sur l'image *Lena* ($D_{\mathsf{xy}} = 10$, $n = 100$ et $\varphi_i = 1$)

distorsion d'insertion soit commune. Dans le premier cas (courbe bleue), σ_{lim} est tel que tous les échantillons sont marqués. On voit qu'il est possible d'obtenir un rapport signal-à-bruit correct même pour des distortions d'attaques très fortes ($n \times E_b/N_0 > 300$ pour $D_{\mathsf{xy}'} = 50$, soit un PSNR de $31,1$ dB). En contrepartie, s'il n'y a pas d'attaque, la performance est à peine meilleure ($n \times E_b/N_0 \simeq 450$ pour $D_{\mathsf{xy}'} = D_{\mathsf{xy}}$). Dans le second cas, la valeur de σ_{lim} fait que seule la moitié des échantillons est marquée. Par conséquent, si cette moitié est effacée (ce qui introduit une distorsion relativement faible), la performance est nulle ($E_b/N_0 = 0$ pour $D_{\mathsf{xy}'} \simeq 30$). Par contre, pour des attaques plus faibles, ce cas est rapidement meilleur que le premier ($n \times E_b/N_0 \simeq 15000$ pour $D_{\mathsf{xy}'} = D_{\mathsf{xy}}$). En fonction de la distorsion d'attaque visée, il faut choisir des paramètres d'insertion différents. Dans cet exemple, si la distorsion d'attaque maximale visée est inférieure à 28, il vaut mieux prendre la seconde solution. Rechercher la meilleure stratégie pour une distorsion d'attaque donnée $D^{\max}_{\mathsf{xy}'}$ consiste à parcourir toutes les valeurs de σ_{lim} possibles, puis à fixer le paramètre λ afin que la distorsion d'insertion soit au niveau voulu et à noter la valeur de E_b/N_0 correspondant à $D^{\max}_{\mathsf{xy}'}$. La stratégie donnant le meilleur résultat est la stratégie optimale.

De plus, en injectant les paramètres de la stratégie optimale dans la formule de l'estimateur (équation 3.2), on remarque une nouvelle forme d'estimation, donnée par

$$\widehat{b_j} = \sum_{i=1}^{m} \varphi_i \times \mathsf{G}(i,j) \times y'_i. \tag{3.18}$$

Cette formule est valable sur l'ensemble des échantillons marqués et dans le cas où l'attaque subie est l'attaque optimale. L'estimation des bits insérés peut donc être faite sans connaissance des paramètres d'insertion (hormis la valeur de σ_{lim}).

Conclusion

À la fin du chapitre précédent, nous avions vu que le filtre de Wiener était une des stratégies d'attaque optimale, permettant à l'attaquant de restaurer l'image marquée et d'en augmenter la qualité. De plus, ce type de filtrage, modélisé par un coefficient multiplicatif devant la formulation de l'énergie de la marque, ne modifie pas les performances de la transmission. Il était donc tout à fait intuitif de penser utiliser un tel filtrage au moment même de l'insertion. Ce chapitre nous a permis de vérifier analytiquement cette hypothèse, en introduisant un nouveau paramètre $\bar{\gamma}_i$ dans la stratégie de défense. On retrouve des formules d'insertion similaires à celles de Cohen et Lapidoth [CL02] vues dans la section 4.3 de la première partie : la marque doit être ajoutée, puis filtrée afin de réduire la distorsion d'insertion.

Appliquer ce filtre modifie la formulation de la distorsion d'insertion D_{xy}. Les règles du jeu sont donc modifiées. Cela modifie la répartition de l'énergie de la marque. On voit apparaître une coupure nette, laissant de côté des échantillons de plus forte énergie.

Chapitre 4

Résultats

Par l'utilisation d'une optimisation max-min, nous avons défini une attaque optimale quelle que soit la stratégie d'insertion choisie. Nous avons également exhibé la répartition de l'énergie de la marque donnant les meilleures performances. Nous allons confronter dans ce chapitre ces stratégies face à des attaques et des techniques d'insertion classiques.

4.1 Évaluation de l'attaque optimale

Afin de mesurer la performance de l'attaque optimale définie dans la section 2.3, nous l'avons appliquée sur des stratégies d'insertion inspirées de l'état de l'art, listées ici :

- la stratégie constante, où l'énergie de la marque est identique pour tous les échantillons du signal à tatouer. D'après la mesure de distorsion utilisée (voir l'équation 2.5), nous avons donc $\sigma^2_{W_i} = D_{\mathsf{xy}}$,
- une stratégie similaire à celle utilisée par Piva *et al.* [PBBC97, PBBC98] et respectant le PSC de Su *et al.* [SG99, SKH02], où l'énergie du $i^{\text{ème}}$ élément de la marque est proportionnelle à la valeur absolue du $i^{\text{ème}}$ échantillon du signal hôte : $\sigma_{W_i} \propto \sigma_{X_i}$.

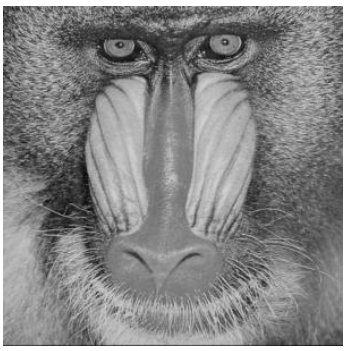
Le signal hôte x est obtenu par la décomposition en ondelettes sur trois niveaux d'une image en niveaux de gris. La sous-bande de plus basse fréquence n'est pas utilisée (figure 4.2) car nos expérimentations montrent que la moindre modification sur ses coefficients est très visible, même si une pondération perceptuelle telle que celle de Watson[1] est utilisée. Les images testées sont données par la figure 4.1. Ce sont des images classiquement utilisées en traitement d'images, de taille 512×512. L'image *Rose* a la particularité de posséder très peu de hautes fréquences, à l'opposé de *Baboon*. Les images *Lena* et *Paper* ont une distribution fréquentielle moyenne. Les variances locales $\sigma^2_{X_i}$ sont calculées sur une fenêtre 5×5 centrée sur l'échantillon considéré. Nous n'avons pas utilisé de pondération perceptuelle ($\varphi_i = 1$), les distorsions d'insertion et d'attaque sont donc des EQM.

[1] Voir la section 2.3.2 de la première partie.

(a) Image *Lena*

(b) Image *Paper. Copyright photo courtesy of Karel de Gendre*

(c) Image *Rose. Copyright photo courtesy of Toni Lankerd*



(d) Image *Baboon*

FIG. 4.1 – Images utilisées : taille 512×512, en niveaux de gris

L'attaque optimale est comparée à deux autres attaques. La première consiste à ajouter un bruit gaussien i.i.d. sur l'ensemble du signal marqué. Cette stratégie est résumée par $\mathsf{e}(i) = (\bar{\gamma}_i = 1, \sigma_{Z_i} = \sqrt{D_{\mathsf{xy}'} - D_{\mathsf{xy}}})$. La seconde est composée d'un filtrage de Wiener et d'un bruit gaussien proportionnel à l'énergie de la marque : $\mathsf{e}(i) = (\gamma_i^{\mathrm{W}}, \sigma_{Z_i} \propto \sigma_{W_i})$. Le facteur de proportionnalité est fixé en respectant la distorsion d'attaque que l'on souhaite atteindre. Les performances sont évaluées en mesurant le rapport signal-à-bruit E_b/N_0, défini par

$$\frac{E_b}{N_0} = \sum_{i=1}^{m} \frac{\gamma_i^2 \sigma_{W_i}^2}{\gamma_i^2 \left(\sigma_{X_i}^2 + \sigma_{W_i}^2\right) + \sigma_{Z_i}^2}. \tag{4.1}$$

Cette formule diffère de l'équation 3.1 car le rapport $1/n$ a été supprimé. De ce fait, les performances données ici correspondent au rapport signal-à-bruit global. Si n bits sont insérés, il doit être divisé par n.

La figure 4.3 montre l'impact des attaques sur un schéma de tatouage avec une répartition uniforme de l'énergie de la marque. Chaque image est marquée en respectant une distorsion d'insertion égale à 10 (soit un PSNR de 38, 13 dB). Elle est ensuite attaquée en faisant varier la force des attaques testées (c'est-à-dire la distorsion $D_{\mathsf{xy}'}$ introduite, reportée en abscisse sur les graphiques). À chaque étape, le rapport E_b/N_0 est calculé (en ordonnée). Plus ce rapport est faible et plus l'attaque est efficace. On peut obtenir depuis cette mesure la capacité totale maximale du canal de tatouage[2]. Nous l'estimons par

$$\mathcal{C} = \frac{m}{2} \log_2 \left[1 + \frac{E_b}{m \times N_0}\right]. \tag{4.2}$$

On voit la nette supériorité de l'attaque optimale sur les deux autres attaques testées. Sur l'image *Rose*, de nombreux échantillons sont annulés, faisant baisser très rapidement la performance de la transmission, jusqu'à l'annulation complète du signal à $D_{\mathsf{xy}'} \simeq 41$. La figure 4.4 reprend le même type de tests, appliqué à un marquage *à la* Piva. On retrouve le bon comportement de l'attaque optimale.

4.2 Comportement de la défense

La stratégie d'attaque définie dans la première section de cette partie donne les résultats espérés : pour une distorsion d'attaque donnée, elle donne le rapport signal-à-bruit le plus faible quelle que soit la stratégie d'insertion adoptée. Nous allons voir dans cette section si la défense basée sur un tatouage additif filtré (avec filtrage de Wiener lors de l'insertion) donne de meilleurs résultats que les répartitions uniformes ou *à la* Piva, et comment elle se comporte face aux attaques présentées au-dessus.

4.2.1 Face à l'attaque optimale

La figure 4.5 donne le comportement des trois stratégies d'insertion face à l'attaque optimale. Pour chaque $D_{\mathsf{xy}'}$, nous calculons la stratégie d'insertion correspondante,

[2]Voir la section 1.4 pour plus de détails.

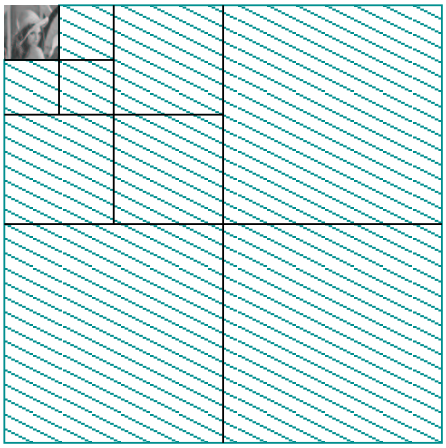


FIG. 4.2 – Ensemble des coefficients formant le signal hôte x (présenté ici avec une DWT sur trois niveaux) : la sous-bande de plus basse fréquence n'est pas utilisée

c'est-à-dire le couple (λ, χ) optimisé pour la distorsion de l'attaque appliquée. Comme prévu, la stratégie définie par la théorie du jeu donne les meilleures performances. Ainsi pour l'image *Paper* avec une distorsion $D_{\mathsf{xy}'} = 30$ (PSNR de $33,4$ dB), l'insertion constante peut atteindre une capacité de 72 bits, l'insertion *à la* Piva 509 bits et notre insertion 2152 bits.

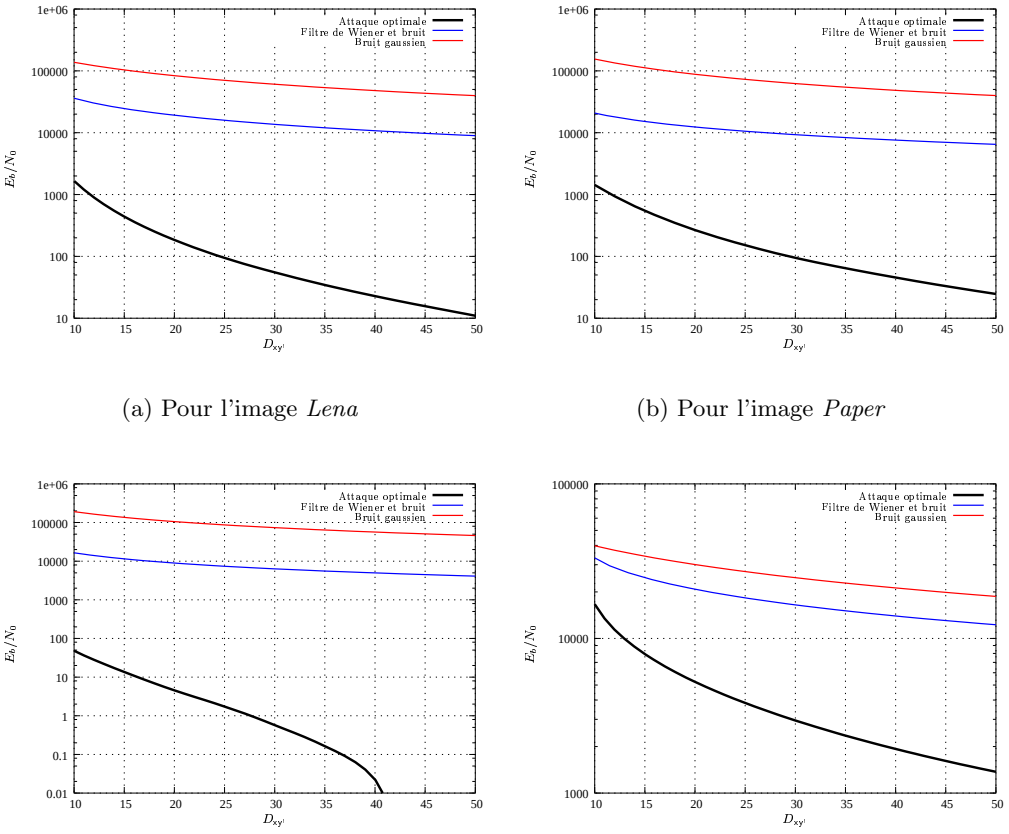
4.2.2 Face à des attaques quelconques

Enfin, afin de valider notre technique d'insertion, nous avons repris les attaques testées dans la première section de ce chapitre, c'est-à-dire l'ajout de bruit gaussien uniforme et l'ajout d'un bruit gaussien proportionnel à l'énergie de la marque[3]. Cette fois aussi, nous ajustons notre stratégie d'insertion en fonction du niveau d'attaque visé. La figure 4.6 montre les résultats obtenus sur les quatre images de test. On voit que si l'attaque optimale n'est pas appliquée, la performance de transmission est meilleure.

Nous avons également appliqué les traitements du test Stirmark [Pet00, sti] sur notre stratégie d'insertion. Nous n'avons pas pris en compte les attaques de type géométrique, et donc seules les traitements suivants ont été testés :

- filtres médians de taille 2×2, 3×3 et 4×4,
- filtre gaussien 3×3,
- attaque FMLR (*frequency mode Laplacian removal*),
- *sharpenning* 3×3,
- compression JPEG avec des qualités allant de 95 % (très bonne qualité) à 10 % (qualité médiocre).

[3]Dans la première section, cette attaque était précédée d'un filtrage de Wiener, mais comme il est préalablement appliqué par notre technique d'insertion, il devient inutile.



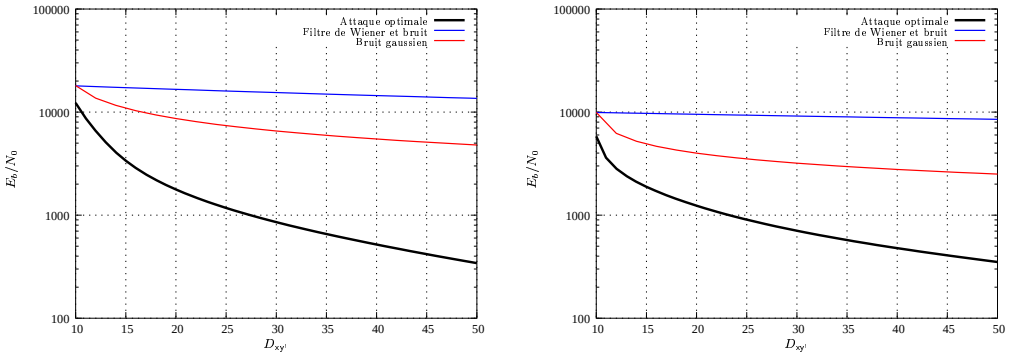
(a) Pour l'image *Lena*

(b) Pour l'image *Paper*

(c) Pour l'image *Rose*

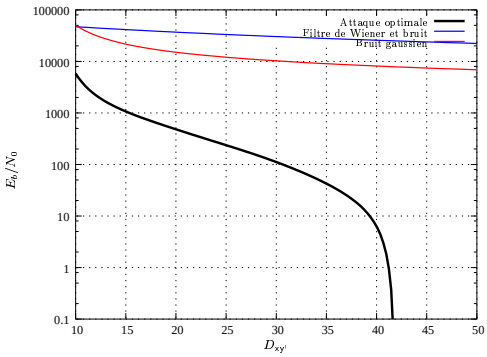
(d) Pour l'image *Baboon*

FIG. 4.3 – Impact de différentes attaques, en utilisant une énergie de marque constante (répartition uniforme) telle que $D_{\mathsf{xy}} = 10$ ($\varphi_i = 1$)

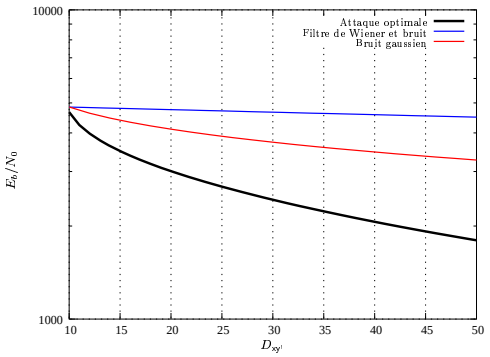


(a) Pour l'image *Lena*

(b) Pour l'image *Paper*

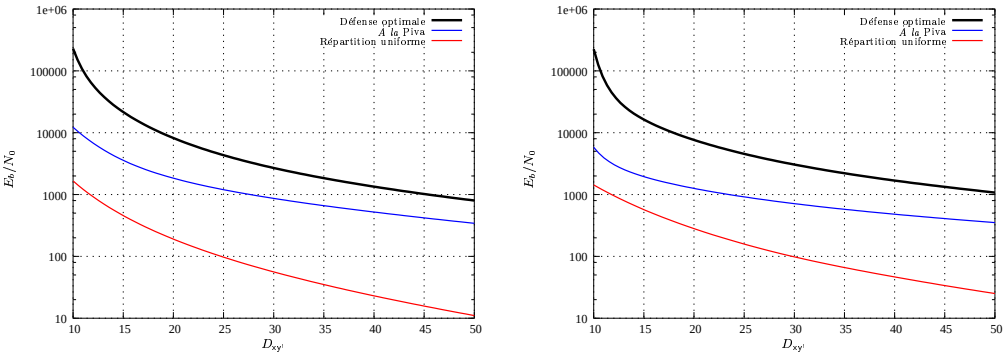


(c) Pour l'image *Rose*



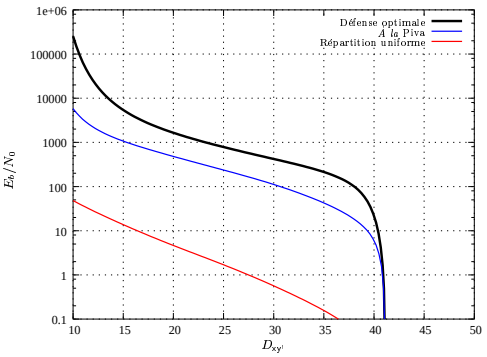
(d) Pour l'image *Baboon*

FIG. 4.4 – Impact de différentes attaques, en utilisant une énergie de marque du type $\sigma_{W_i} \propto \sigma_{X_i}$ telle que $D_{\mathsf{xy}} = 10$ ($\varphi_i = 1$)

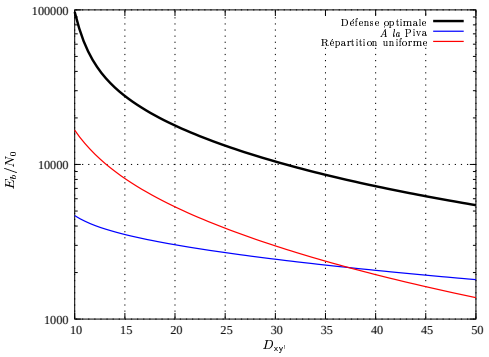


(a) Pour l'image *Lena*

(b) Pour l'image *Paper*



(c) Pour l'image *Rose*



(d) Pour l'image *Baboon*

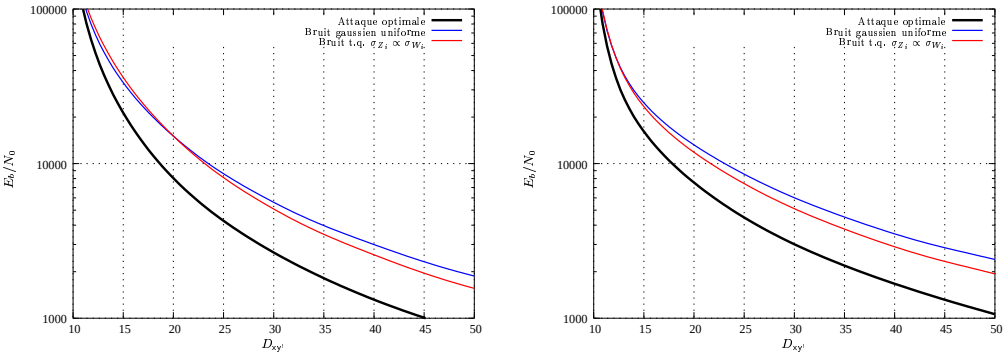
FIG. 4.5 – Performances de plusieurs répartitions de l'énergie de la marque face à l'attaque optimale. La distorsion d'insertion est fixée à $D_{\mathsf{xy}} = 10$ ($\varphi_i = 1$)

Comme nous ne pouvions prévoir à l'avance la distorsion introduite par ces attaques et donc ajuster au mieux la stratégie d'insertion, nous avons fixé les paramètres λ et χ afin de viser une attaque de distorsion $D_{\mathsf{xy}'} = 40$ (PSNR de 32 dB). La figure 4.7 indique les résultats. Les losanges rouges correspondent aux filtrage médians, les verts à l'attaque FMLR, les noirs au filtre gaussien et les bleus au *sharpenning* [4]. Malgré le fait que la stratégie d'insertion ne soit pas spécifiquement adaptée à chacune des attaques, les performances restent excellentes. À titre de comparaison, les résultats obtenus face à l'attaque optimale ont été ajoutés (courbe en trait gras). On voit que les attaques de Stirmark sont bien moins performantes que l'attaque que nous avons définie. En particulier, alors que la compression JPEG est souvent utilisée pour tester la performance des schémas de tatouage, on remarque que son impact sur les performances est assez faible.

Conclusion

Ces premiers résultats sont tout à fait conformes à la théorie du jeu utilisée pour notre optimisation. Quelle que soit la technique d'insertion utilisée, l'attaque optimale est la meilleure (celle qui fait baisser le plus le rapport signal-à-bruit de la transmission). La stratégie d'insertion correspondante, s'auto-appliquant un filtrage de Wiener, est la meilleure stratégie face à cette attaque. Si une attaque autre que l'optimale est alors appliquée, les résultats sont encore améliorés.

[4] Ce dernier point n'apparaît que pour l'image *Rose* car il engendre énormément de distorsion. Néanmoins, la perte de performance est très faible (par exemple, sur *Baboon*, on obtient $E_b/N_0 > 11000$ pour $D_{\mathsf{xy}'} > 2000$).



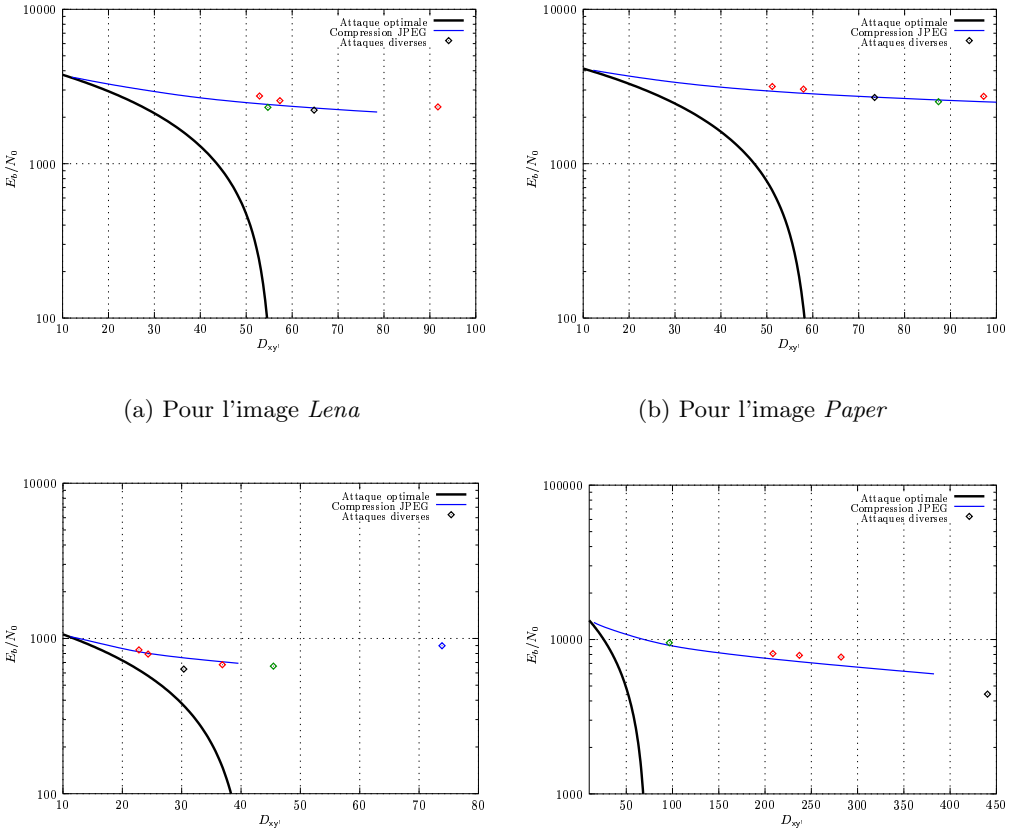
(a) Pour l'image *Lena*

(b) Pour l'image *Paper*

(c) Pour l'image *Rose*

(d) Pour l'image *Baboon*

FIG. 4.6 – Impact de l'ajout de bruit gaussien sur la performance de la défense optimale. La distorsion d'insertion est fixée à $D_{\mathsf{xy}} = 10$ ($\varphi_i = 1$)



(a) Pour l'image *Lena*

(b) Pour l'image *Paper*

(c) Pour l'image *Rose*

(d) Pour l'image *Baboon*

FIG. 4.7 – Impact des attaques implémentées par Stirmark sur la performance de la défense optimale. La distorsion d'insertion est fixée à $D_{\mathsf{xy}} = 10$ et nous visons une attaque de distorsion $D_{\mathsf{xy}'} = 40$ ($\varphi_i = 1$)

Troisième partie

Prise en compte de l'information adjacente

Introduction

L'étude de l'existant de la première partie a montré que le tatouage pouvait s'assimiler à un problème de communication. Nous avons vu également qu'un type de canal était particulièrement adapté au tatouage : le canal avec information adjacente disponible à l'encodeur (*side information*). Dans ce cas, le message que l'on souhaite transmettre (contraint en énergie) est bruité par deux éléments : un bruit déjà présent et connu lors de l'envoi des données et un bruit ajouté lors de la transmission (figure 8). Le schéma de Costa [Cos83] (ICS pour *ideal Costa scheme*), applicable à des signaux i.i.d. gaussiens, montre qu'il est possible d'obtenir de meilleures performances par rapport à un canal AWGN classique (voir le chapitre 3 de la première partie).

La transmission d'un message par tatouage est typiquement un problème de transmission sur un canal avec information adjacente. Dans les schémas de tatouage additif, le signal hôte est un bruit venant perturber la transmission de la marque et il est parfaitement connu avant le processus de codage. La prise en compte de cette information adjacente pourrait donc permettre un fort gain de performance, comme le montre la figure 9. Néanmoins, il est impossible d'appliquer le schéma de Costa tel quel sur notre technique. D'une part, il s'appuie sur la construction d'un dictionnaire aléatoire de grande dimension, impossible à implémenter directement. D'autre part, il suppose des signaux gaussiens et i.i.d. (pour l'information adjacente et le bruit ajouté lors de la transmission), ce qui est en contradiction avec nos hypothèses de départ.

Cette troisième partie va montrer comment adapter le schéma de Costa à notre technique de tatouage par étalement de spectre et optimisation grâce à la théorie des jeux. Nous allons d'abord montrer que l'étalement de spectre définit un sous-espace linéaire correspondant aux distributions statistiques imposées par Costa, et nous verrons comment optimiser ses paramères (par max-min). La section suivante exposera la construction d'un dictionnaire structuré issu de codes correcteurs d'erreur, adapté à l'ICS. Nous introduirons dans le chapitre 2.4 une technique permettant de supprimer l'interférence inter-symboles (un facteur limitant de l'étalement de spectre). Enfin, le chapitre 3 étudiera deux extensions de notre schéma : l'introduction de désynchronisations dans la mesure de performance, et la prise en compte de la réalisation du signal (afin d'améliorer notre modèle statistique).

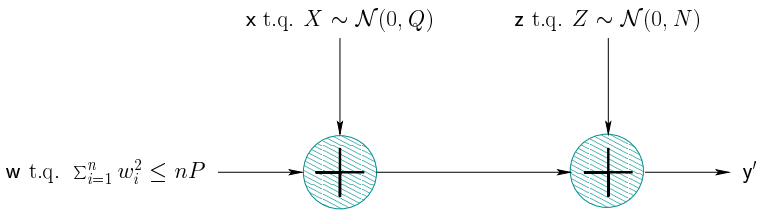


FIG. 8 – Transmission d'un signal w *via* un canal gaussien avec information adjacente disponible à l'encodage

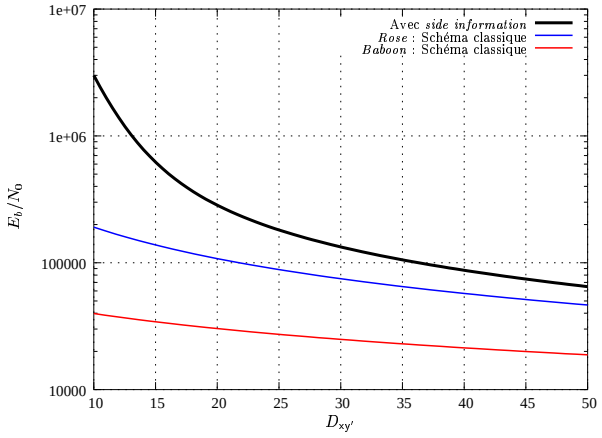


FIG. 9 – Performances possibles avec suppression de l'interférence du signal hôte grâce à la prise en compte de l'information adjacente (énergie d'insertion uniforme, attaque par bruit gaussien uniforme, $D_{\mathsf{xy}} = 10$, $\varphi_i = 1$ et $n = 100$)

Chapitre 1

Étalement de spectre et information adjacente

Les signaux mutimedia suivent rarement une loi Normale identiquement distribuée. Nous avons alors considéré un modèle plus général pour la construction du notre schéma : le signal hôte x, la marque w et le bruit d'attaque z sont des signaux gaussiens non i.i.d. modélisés par des ensembles de variables aléatoires. Or, une des hypothèses de la démonstration de Costa est que les signaux soient i.i.d. et gaussiens. Son schéma n'est donc pas applicable directement. Une solution est donnée par les canaux parallèles gaussiens déjà évoqués dans l'introduction de la seconde partie, mais ils ne permettent pas de définir un équilibre global du jeu (sur l'ensemble des canaux) par optimisation max-min.

L'étalement de spectre peut être vu comme la projection d'un signal sur un ensemble de porteuses, définissant un sous-espace linéraire. Les signaux ainsi projetés suivent une loi Normale, respectant la condition de l'ICS. Nous redéfinissons dans ce chapitre les paramètres de l'étalement de spectre vu dans la partie 2 (estimateur optimal et jeu entre attaquant et défenseur) afin d'appliquer l'ICS dans le sous-espace obtenu.

1.1 Fonctions de projection

Les fonctions de projection (*spread transforms*) définissent un sous-espace linéaire par la projection des éléments du signal considéré sur un ensemble de porteuses. On passe d'un signal de dimension m à sa version projetée de dimension n.

Le tatouage additif filtré (plus performant que le simple tatouage additif, comme vu dans le chapitre 3 de la partie précédente) utilisant la technique de l'étalement de spectre est défini par

$$\begin{aligned} y_i &= \bar{\gamma}_i \left[x_i + w_i \right] \\ &= \bar{\gamma}_i \left[x_i + \frac{\sigma_{W_i}}{\sqrt{n}} \sum_{j=1}^{n} \mathsf{G}(i,j) \times b_j \right], \end{aligned} \tag{1.1}$$

avec $\mathsf{G} \in \mathbb{R}^{n.m}$ matrice pseudo-aléatoire, telle que $\mathbb{E}\left[\mathsf{G}(i,j)\right] = 0$ et $\mathbb{E}\left[\mathsf{G}(i,j)^2\right] = 1$, définissant l'ensemble des n porteuses et b la suite de symboles (d'énergie unitaire) que l'on souhaite insérer. L'intérêt d'un facteur multiplicatif $\bar{\gamma}_i$ à l'insertion a été démontré dans la partie 2 : il permet de réduire la distorsion d'insertion (par un filtre de Wiener) sans modifier la performance du schéma de tatouage. Nous avons modélisé les attaques sous la forme SAWGN : $y'_i = \bar{\bar{\gamma}}_i \times y_i + z_i$, avec z bruit non i.i.d. gaussien. Ce type de modélisation convient à de nombreuses attaques (compression avec perte, bruit corrélé, filtrage, ...). Nous reprenons la notation $\gamma_i = \bar{\gamma}_i \times \bar{\bar{\gamma}}_i$. Nous avons vu dans la partie précédente que l'extraction depuis le signal y' se fait par

$$\widehat{b_j} = \sum_{i=1}^{m} \beta_i \times \mathsf{G}(i,j) \times y'_i, \tag{1.2}$$

avec β_i défini par l'équation 3.2 de la page 84. L'équation ci-dessus peut être vue comme la projection du signal y' sur les porteuses définies par G. De même, on peut considérer b comme étant le signal de la marque dans le sous-espace. L'étalement de spectre correspond donc à l'ajout de deux signaux dans un sous-espace linéraire obtenu par projection. Les projetés de x, y' et z sont respectivement définis par

$$x_j^{\mathrm{st}} = \sum_{i=1}^{m} \beta_i\gamma_i \times \mathsf{G}(i,j) \times x_i \tag{1.3}$$

$$y'^{\mathrm{st}}_j = \sum_{i=1}^{m} \beta_i \times \mathsf{G}(i,j) \times y'_i \tag{1.4}$$

$$z_j^{\mathrm{st}} = \sum_{i=1}^{m} \beta_i \times \mathsf{G}(i,j) \times z_i \tag{1.5}$$

et suivent une loi Normale uniforme (voir la section 1.3 de la partie 2). Dans le sous-espace, l'insertion donnée par l'équation 1.1 correspond à[1]

$$\mathsf{y}^{\mathrm{st}} = \mathsf{x}^{\mathrm{st}} + \mathsf{w}^{\mathrm{st}}. \tag{1.6}$$

L'hypothèse de Costa sur la distribution des signaux est donc vérifiée et son schéma peut être appliqué dans le sous-espace. En suivant ce paradigme, le vecteur b n'est plus une suite de symboles mais doit être défini tel que décrit par l'ICS, c'est-à-dire par $\mathsf{b} = \mathsf{u}^\star - \alpha\mathsf{x}^{\mathrm{st}}$, avec $\mathsf{u}^\star$ mot de code issu d'un dictionnaire structuré. Par souci de cohérence, nous notons ce signal de marque w^{st}. La formule de l'insertion est alors

$$y_i = \bar{\gamma}_i \left[x_i + \frac{\sigma_{W_i}}{\|\mathsf{w}^{\mathrm{st}}\|} \sum_{j=1}^{n} \mathsf{G}(i,j) \times w_j^{\mathrm{st}} \right]. \tag{1.7}$$

[1] L'égalité n'est pas tout à fait vraie car il faut ajouter l'interférence introduite par la non-orthogonalité des porteuses (interférence inter-symboles). Néanmoins, nous la négligeons pour la suite de ce chapitre. Une solution permettant de la supprimer sera étudiée dans la section 2.4.

Dans le sous-espace défini par l'étalement de spectre, les signaux sont modélisés par une loi Normale : $X^{\text{st}} \sim \mathcal{N}(0, Q)$ et $Z^{\text{st}} \sim \mathcal{N}(0, N)$, dont les variances respectives sont données par

$$Q = \sum_{i=1}^{m} \beta_i^2 \gamma_i^2 \times \sigma_{X_i}^2 \tag{1.8}$$

$$N = \sum_{i=1}^{m} \beta_i^2 \times \sigma_{Z_i}^2. \tag{1.9}$$

Une des hypothèses du schéma de Costa est la contrainte d'énergie sur le signal transmis w. Ici, elle peut s'écrire $\sum_{j=1}^{n} \left[w_j^{\text{st}}\right]^2 \leq nP$. D'après la formule d'insertion de l'équation 1.1, la valeur de P est

$$P = \frac{1}{n} \left[\sum_{i=1}^{m} \beta_i \gamma_i \times \sigma_{W_i} \right]^2. \tag{1.10}$$

Malgré le fait que $\sigma_{W_i} \ll \sigma_{X_i}$ afin de limiter la distorsion introduite par la marque, l'équation ci-dessous montre que l'énergie répartie sur l'ensemble des échantillons du signal marqué est concentrée dans le sous-espace : la valeur de Q et de N est fonction de m tandis que celle de P est fonction de m^2/n. De ce fait, il est possible d'avoir $P > Q$ pour des rapports m/n importants.

1.2 Estimateur optimal

En prenant en compte l'information adjacente, la capacité qu'il est possible d'atteindre est fonction croissante du rapport P/N. Nous décidons de prendre ce rapport comme mesure de performance de notre schéma de tatouage. Trouver l'estimateur optimal consiste à déterminer la forme de β_i maximisant P/N :

$$\beta_i^\star = \arg\max_{\beta_i} \left\{ \frac{P}{N} = \frac{\left[\sum_{i=1}^{m} \beta_i \gamma_i \times \sigma_{W_i}\right]^2}{n \sum_{i=1}^{m} \beta_i^2 \times \sigma_{Z_i}^2} \right\}. \tag{1.11}$$

En recherchant la valeur de β_i annulant la dérivée de l'équation de P/N, on obtient

$$\beta_i^\star \propto \frac{\gamma_i \sigma_{W_i}}{\sigma_{Z_i}^2}, \tag{1.12}$$

En injectant cette valeur dans l'équation 1.11, le rapport signal-à-bruit est exprimable par

$$\frac{E_b}{N_0} = \frac{P}{N} = \frac{1}{n} \sum_{i=1}^{m} \frac{\gamma_i^2 \sigma_{W_i}^2}{\sigma_{Z_i}^2}. \tag{1.13}$$

Tout comme l'expression trouvée dans la partie concernant le tatouage sans prise en compte de l'information adjacente, cette formule est assez conforme à l'intuition. Le

nominateur correspond à l'énergie de la marque insérée et le dénominateur est l'énergie du bruit d'attaque. En toute logique, et du fait de la prise en compte de l'information adjacente, le signal hôte a disparu de la mesure de performance par rapport à la formule utilisée dans la seconde partie (équation 1.19 de la page 67). Il faut tout de même remarquer que même si nous notons $E_b/N_0 = P/N$ par analogie avec un canal gaussien classique, ce rapport ne correspond pas au véritable rapport signal-à-bruit observé en sortie du canal, donné par

$$\frac{P(P+Q+N)}{N(N+P)}.$$

Une explication plus détaillée sur ce résultat est fourni dans l'annexe B.

On note que la formule optimale $\beta_i^\star$ n'est pas la formule naturelle de la projection inverse[2]. De ce point de vue, l'étalement de spectre tel que nous l'utilisons n'est pas une projection réversible, contrairement aux fonctions de projection utilisées dans d'autres schémas de tatouage [CW00, EBTG02].

1.3 Résolution par la théorie des jeux

La formule du rapport signal-à-bruit de l'équation 1.13 est la mesure de performance que nous allons utiliser pour déterminer la meilleure stratégie de répartition de l'énergie de la marque. Nous utilisons cette fois encore une approche de type max-min. Les distorsions d'insertion et d'attaque restent inchangées par rapport au chapitre 3 de la seconde partie, c'est-à-dire

$$D_{\mathsf{xy}} = \frac{1}{m}\sum_{i=1}^{m}\varphi_i^2\left[\sigma_{X_i}^2\left(1-\bar{\gamma}_i\right)^2+\bar{\gamma}_i^2\sigma_{W_i}^2\right] \tag{1.14}$$

$$D_{\mathsf{xy}'} = \frac{1}{m}\sum_{i=1}^{m}\varphi_i^2\left[\sigma_{X_i}^2\left(1-\gamma_i\right)^2+\gamma_i^2\sigma_{W_i}^2+\sigma_{Z_i}^2\right]. \tag{1.15}$$

1.3.1 Attaque optimale

L'attaque optimale est l'attaque minimisant la performance du schéma et respectant une contrainte de distorsion donnée. Comme précédemment, nous utilisons une formulation lagrangienne :

$$\begin{aligned}\mathsf{a}_{\mathsf{e}}(D_{\mathsf{xy}'}^{\max}) &= \arg\min_{\mathsf{a}\in\mathcal{A}}\left\{\frac{E_b}{N_0}(\mathsf{a},\mathsf{e})+\lambda'\left[D_{\mathsf{xy}'}-D_{\mathsf{xy}'}^{\max}\right]\right\}\\ &= \arg\min_{\mathsf{a}\in\mathcal{A}}\left\{J_\lambda = n\frac{E_b}{N_0}(\mathsf{a},\mathsf{e})+\lambda m\left[D_{\mathsf{xy}'}-D_{\mathsf{xy}'}^{\max}\right]\right\}.\end{aligned} \tag{1.16}$$

La fonctionnelle J_λ étant additive, la minimisation peut se faire terme à terme. Les paramètres d'insertion pour le $i^{\text{ème}}$ échantillon sont donc définis par

$$(\gamma_i^\star,\sigma_{Z_i}^\star) = \arg\min_{\gamma_i,\sigma_{Z_i}\geq 0}\left\{J_\lambda^i = \frac{\gamma_i^2\sigma_{W_i}^2}{\sigma_{Z_i}^2}+\lambda\varphi_i^2\left[\sigma_{X_i}^2\left(1-\gamma_i\right)^2+\gamma_i^2\sigma_{W_i}^2+\sigma_{Z_i}^2\right]\right\}, \tag{1.17}$$

[2]Qui serait $\beta_i \propto \sigma_{W_i}/\sqrt{n}\gamma_i$.

avec $\gamma_i = \bar{\gamma}_i \times \bar{\bar{\gamma}}_i$. Nous allons d'abord rechercher une solution sur $\mathbb{R} \times \mathbb{R}^{+*}$, puis nous explorerons les cas limites (c'est-à-dire $\sigma_{Z_i} = 0$). Les dérivées de J^i_λ par rapport à γ_i et à $\sigma^2_{Z_i}$ sont données respectivement par

$$\frac{\partial J^i_\lambda}{\partial \gamma_i} = \frac{2\gamma_i \sigma^2_{w_i}}{\sigma^2_{Z_i}} + 2\lambda\varphi_i^2 \left[\gamma_i \left[\sigma^2_{X_i} + \sigma^2_{W_i}\right] - \sigma^2_{X_i}\right] \tag{1.18}$$

$$\frac{\partial J^i_\lambda}{\partial \sigma^2_{Z_i}} = -\frac{\gamma_i^2 \sigma^2_{W_i}}{\sigma^4_{Z_i}} + \lambda\varphi_i^2. \tag{1.19}$$

Leur annulation commune donne les valeurs

$$\gamma_i^{\mathrm{a}} = \gamma_i^{\mathrm{w}} - \frac{\sigma_{W_i}}{\sqrt{\lambda}\varphi_i \left(\sigma^2_{X_i} + \sigma^2_{W_i}\right)} \tag{1.20}$$

$$\sigma^{\mathrm{a}}_{Z_i} = \sqrt{\gamma_i^{\mathrm{a}} \left(\gamma_i^{\mathrm{w}} - \gamma_i^{\mathrm{a}}\right) \left(\sigma^2_{X_i} + \sigma^2_{W_i}\right)}, \tag{1.21}$$

avec γ_i^{w} le facteur d'atténuation correspondant à un filtre de Wiener à l'insertion. La formule de $\sigma^{\mathrm{a}}_{Z_i}$ est identique à celle trouvée dans la seconde partie de ce manuscrit, où l'information adjacente n'était pas prise en compte. La formule de γ_i^{a} correspond au facteur multiplicatif d'un filtre de Wiener dont le but serait d'atténuer la marque et le bruit $z_i/\gamma_i^{\mathrm{a}}$ (voir l'équation 2.19 de la page 78). Ces deux paramètres définissent la stratégie d'attaque notée $\mathsf{a_I}(i) = (\gamma_i^{\mathrm{a}}/\bar{\gamma}_i, \sigma^{\mathrm{a}}_{Z_i})$. Ces optimaux locaux ne sont pas valides dans tous les cas. Du fait que σ_{Z_i} soit obligatoirement supérieur ou égal à zéro, les résultats ci-dessus ne sont valables que si la contrainte suivante est respectée :

$$\begin{aligned} \sigma^{\mathrm{a}}_{Z_i} \geq 0 &\Leftrightarrow \gamma_i^{\mathrm{a}} \geq 0 \\ &\Leftrightarrow \sigma_{W_i} \leq \sqrt{\lambda}\varphi_i \sigma^2_{X_i}. \end{aligned} \tag{1.22}$$

Le domaine respectant cette contrainte est noté $\mathcal{D}_2$ et celui ne la respectant pas est noté $\mathcal{D}_1$. Contrairement au cas traité dans la partie 2, nous n'avons ici que deux domaines, car la contrainte $\bar{\gamma}_i \leq \gamma_i^{\mathrm{w}}$ est toujours vraie.

Considérons maintenant le bord du domaine de validité $\sigma_{Z_i} = 0$. La valeur minimisant la fonctionnelle J^i_λ est $\bar{\bar{\gamma}}_i = 0$. Cela définit la stratégie $\mathsf{a_E}(i) = (0, 0)$. Nous avons donc deux attaques possibles à associer à deux domaines. L'annexe A.2.2 nous indique que la stratégie $\mathsf{a_E}(i)$ est optimale sur $\mathcal{D}_1$ et que la stratégie $\mathsf{a_I}(i)$ est optimale sur $\mathcal{D}_2$.

1.3.2 Défense

Afin de trouver la défense optimale (c'est-à-dire les couples $\mathsf{e}(i) = (\bar{\gamma}_i, \sigma_{W_i})$) face à l'attaque que nous venons de définir, nous considérons une optimisation de type max-min, appliquée à une formulation lagrangienne :

$$\begin{aligned} \mathsf{e}(D^{\max}_{\mathsf{xy}}, D^{\max}_{\mathsf{xy'}}) &= \arg\max_{\mathsf{e}\in\mathcal{E}} \left\{ J_\lambda - \chi \left[D_{\mathsf{xy}} - D^{\max}_{\mathsf{xy}}\right] \right\} \\ &= \arg\max_{\mathsf{e}\in\mathcal{E}} \left\{ J_\chi = J_\lambda - \chi m \left[D_{\mathsf{xy}} - D^{\max}_{\mathsf{xy}}\right] \right\}. \end{aligned} \tag{1.23}$$

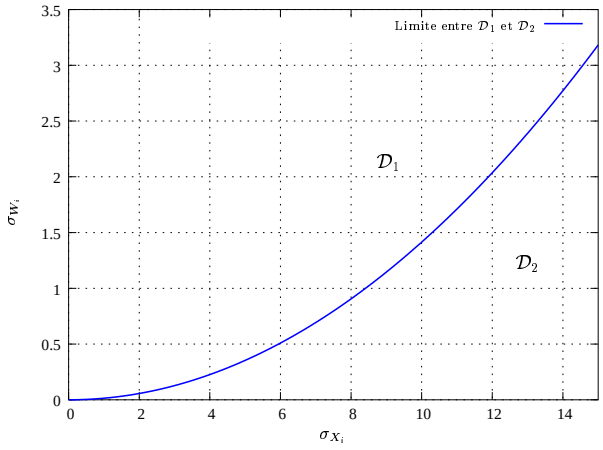


FIG. 1.1 – Les deux domaines d'attaque définis par la contrainte $\sigma_{Z_i} \geq 0$ ($\lambda = 0,02$ et $\varphi_i = 1$)

Et comme J_χ est une fonctionnelle additive séparable, nous pouvons optimiser échantillon par échantillon :

$$(\bar{\gamma}_i^\star, \sigma_{W_i}^\star) = \arg \max_{\bar{\gamma}_i, \sigma_{W_i} \geq 0} \left\{ J_\chi^i = J_\lambda^i - \chi \varphi_i^2 \left[\sigma_{X_i}^2 (1 - \bar{\gamma}_i)^2 + \bar{\gamma}_i^2 \sigma_{W_i}^2 \right] \right\}. \tag{1.24}$$

Si l'on se place sur le domaine $\mathcal{D}_1$, où l'attaque par annulation $\mathsf{a}_\mathsf{E}(i)$ est optimale, la dérivée de J_χ^i suivant σ_{W_i} est toujours négative : la valeur optimale est donc la limite inférieure du domaine, définie par l'équation 1.22. L'annulation de la dérivée de J_χ^i suivant $\bar{\gamma}_i$ nous donne $\bar{\gamma}_i = \gamma_i^{\mathrm{W}}$. La stratégie de défense face à $\mathsf{a}_\mathsf{E}(i)$ est donc $\mathsf{e}_\mathsf{E}(i) = (\bar{\gamma}_i = \gamma_i^{\mathrm{W}}, \sigma_{W_i} = \sqrt{\lambda} \varphi_i \sigma_{X_i}^2)$.

Dans le domaine $\mathcal{D}_2$, l'annulation de la dérivée de J_χ^i suivant $\bar{\gamma}_i$ nous donne cette fois aussi $\bar{\gamma}_i = \gamma_i^{\mathrm{W}}$. La dérivée suivant σ_{W_i} est égale à

$$\frac{\partial J_\chi^i}{\partial \sigma_{W_i}} = 2\sigma_{X_i}^2 \frac{\varphi_i^2 (\lambda - \chi) \sigma_{X_i}^2 \sigma_{W_i} + \sqrt{\lambda} \varphi_i \left(\sigma_{X_i}^2 - \sigma_{W_i}^2 \right) - \sigma_{W_i}}{\left(\sigma_{X_i}^2 + \sigma_{W_i}^2 \right)^2}, \tag{1.25}$$

et sa racine, unique sur $\mathcal{D}_2$, est donnée par

$$\sigma_{W_i}^{\mathrm{a}} = \frac{\varphi_i^2 (\lambda - \chi) \sigma_{X_i}^2 - 1 + \sqrt{\left(\varphi_i^2 (\lambda - \chi) \sigma_{X_i}^2 - 1 \right)^2 + 4\lambda \varphi_i^2 \sigma_{X_i}^2}}{2\sqrt{\lambda} \varphi_i}. \tag{1.26}$$

La stratégie optimale vis-à-vis de $\mathsf{a}_\mathsf{I}(i)$ est donc $\mathsf{e}_\mathsf{I}(i) = (\gamma_i^{\mathrm{W}}, \sigma_{W_i}^{\mathrm{a}})$. De plus, cette stratégie est dans tous les cas meilleure que $\mathsf{e}_\mathsf{E}(i)$: la distorsion d'insertion est plus faible et la fonctionnelle J_λ^i sera plus élevée (voir l'équation A.16 de la page 174).

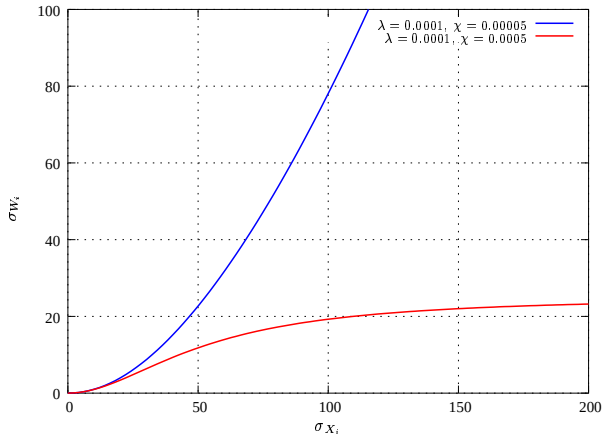


FIG. 1.2 – Deux stratégies d'insertion optimales ($\varphi_i = 1$)

1.3.3 Remarques

Tout comme la stratégie d'insertion définie dans la section 3 de la partie précédente, nous avons montré que la défense a tout intérêt à s'auto-appliquer un filtrage de Wiener après insertion de la marque. La figure 1.2 montre les stratégies d'insertion obtenues pour deux couples (λ, χ) différents. Contrairement aux stratégies vues dans la partie 2, tous les échantillons doivent être marqués. La force d'insertion est croissante en φ_i : les échantillons perceptuellement importants seront plus marqués. On voit aussi que suivant le signe de $\lambda - \chi$, la forme de l'insertion tend soit vers une constante

$$\lim_{\sigma_{X_i} \to \infty} \sigma_{W_i} = \frac{\sqrt{\lambda}}{\varphi_i(\lambda - \chi)}$$

soit vers une forme parabolique

$$\lim_{\sigma_{X_i} \to \infty} \sigma_{W_i} = \varphi_i \frac{\lambda - \chi}{\sqrt{\lambda}} \sigma_{X_i}^2.$$

Lorsque l'attaque optimale est appliquée, seule la stratégie $\mathsf{a_I}$ sera utilisée. En injectant les paramètres de cette stratégie et ceux de la défense dans la formule de l'estimateur optimal défini dans la section 1.2, on trouve $\beta_i^\star \propto \varphi_i$. Cette propriété avait déjà été montrée pour l'estimateur de la seconde partie. Néanmoins, elle n'était applicable que pour les échantillons marqués (il fallait donc connaître la limite entre échantillons marqués et non marqués). Or ici, cet estimateur est utilisable sur l'ensemble du signal y' reçu. Il n'est donc pas nécessaire de connaître les paramètres d'insertion utilisés pour estimer la marque de façon optimale. L'extraction d'un message est totalement aveugle.

1.4 Résultats

Afin de tester la pertinence de notre schéma, nous l'avons comparé à des attaques et des défenses (répartitions de l'énergie de la marque) classiques. Les performances sont quantifiées par le rapport signal-à-bruit global, donné par

$$\frac{E_b}{N_0} = \sum_{i=1}^{m} \frac{\gamma_i^2 \sigma_{W_i}^2}{\sigma_{Z_i}^2}. \tag{1.27}$$

Ce rapport est à diviser par la dimension n du sous-espace utilisé (correspondant à la dimension du signal w^{st} que l'on souhaite transmettre). En utilisant un schéma exploitant l'information adjacente disponible à l'encodeur, la capacité totale maximale est estimée par

$$\mathcal{C} = \frac{m}{2} \log_2 \left[1 + \frac{E_b}{m \times N_0} \right]. \tag{1.28}$$

Le signal hôte est obtenu par la décomposition en ondelettes sur trois niveaux de quatre images de test (présentées par la figure 4.1 de la page 92). Ce sont des images en niveaux de gris de taille 512×512. Comme pour les résultats du chapitre 4 (partie 2), la sous-bande de plus basse fréquence n'est pas utilisée.

1.4.1 Comportement de l'attaque optimale

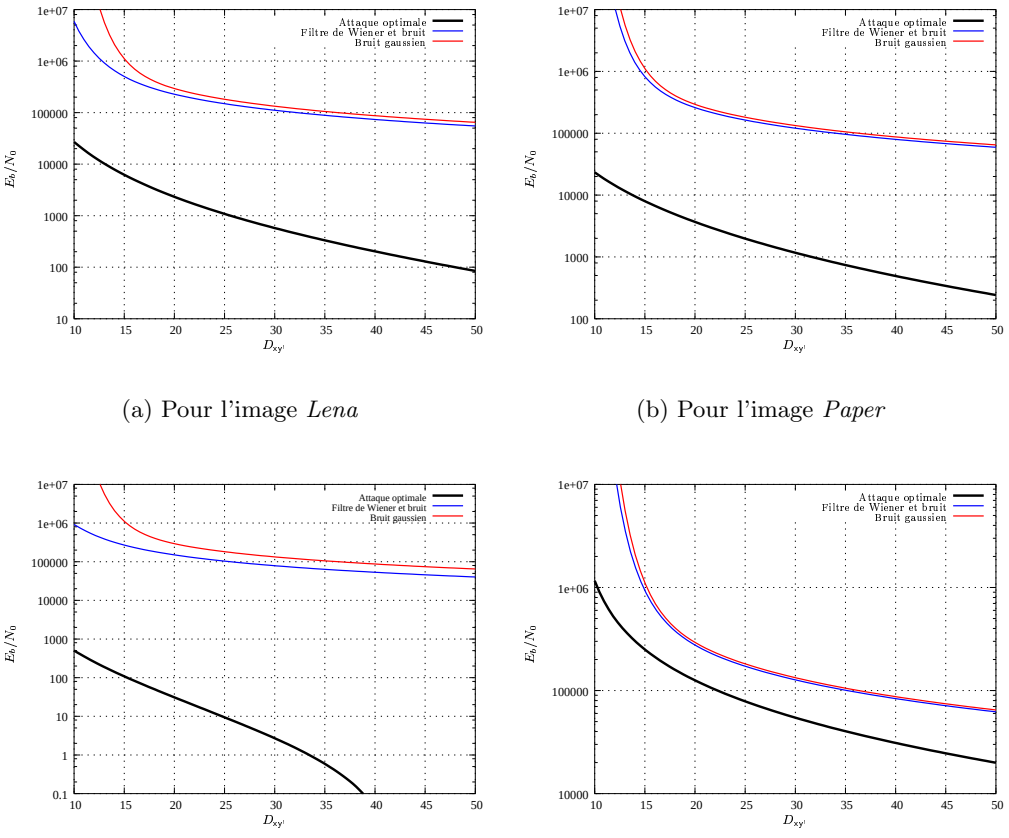
Les attaques testées sont les mêmes que celles du chapitre 4 de la seconde partie : l'ajout d'un bruit gaussien uniforme ($\mathsf{a}(i) = (\bar{\gamma}_i = 1, \sigma_{W_i} = \sqrt{D_{\mathsf{xy}'} - D_{\mathsf{xy}}})$), une combinaison de filtrage de Wiener (afin d'atténuer le bruit de la marque et de restaurer le signal hôte d'origine) et de bruit gaussien proportionnel à l'énergie de la marque en chacun des échantillons ($\mathsf{a}(i) = (\gamma_i^{\mathrm{W}}, \sigma_{Z_i} \propto \sigma_{W_i})$) et bien sûr l'attaque optimale définie dans la section 1.3.1 de ce chapitre.

Ces attaques sont appliquées à des stratégies d'insertion : insertion uniforme (énergie de la marque constante en chacun des échantillons) et insertion avec énergie de la marque proportionnelle à l'énergie de l'échantillon considéré (similaire à la technique de Piva et respectant le PSC de Su *et al.*). La distorsion d'insertion est fixée à 10 (PSNR de $38,13$ dB).

Les figures 1.3 et 1.4 donnent les résultats de ce premier test. La distorsion d'attaque est reportée en abscisse et la performance E_b/N_0, calculée par la formule de l'équation 1.27, est en ordonnée. La baisse de performance introduite par l'attaque optimale est bien plus importante qu'avec les deux autres attaques testées, et ce quelle que soit la technique d'insertion choisie et l'image hôte utilisée. Cela confirme l'intérêt de l'attaque SAWGN.

1.4.2 Performances de la défense

Nous testons maintenant les performances de la défense définie dans la section précédente. La figure 1.5 montre les résultats obtenus pour les trois stratégies d'insertion



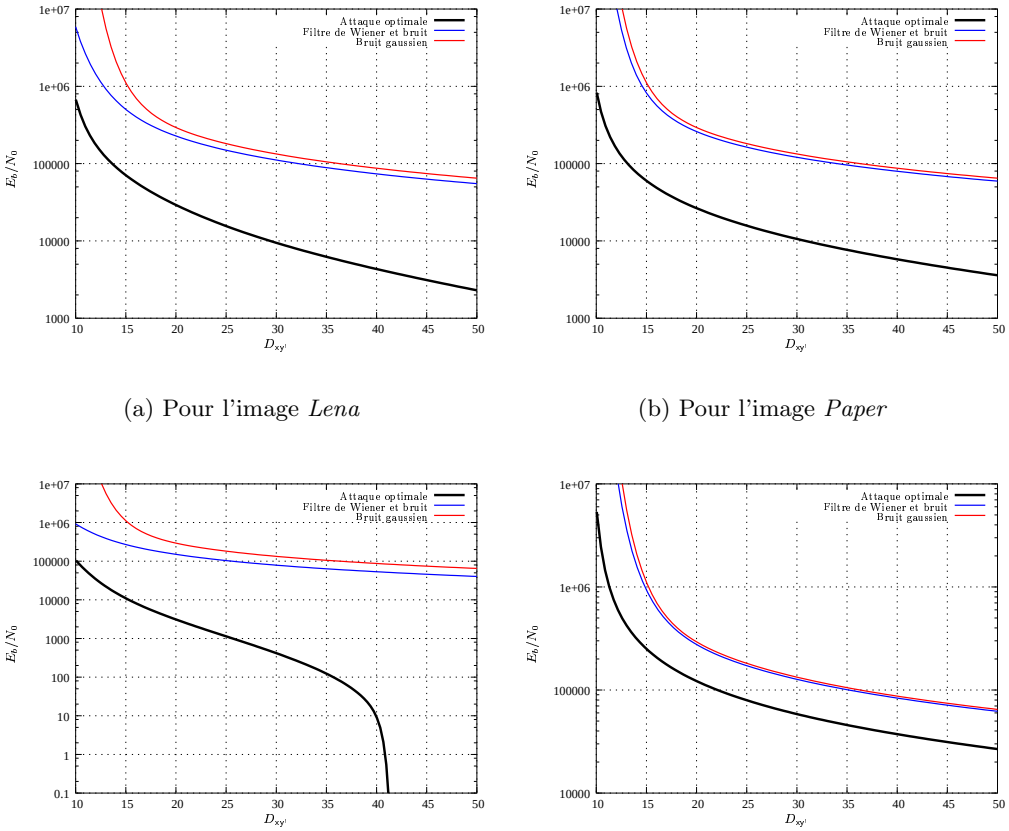
(a) Pour l'image *Lena*

(b) Pour l'image *Paper*

(c) Pour l'image *Rose*

(d) Pour l'image *Baboon*

FIG. 1.3 – Impact de différentes attaques sur le tatouage avec information adjacente, en utilisant une énergie de marque constante (répartition uniforme) telle que $D_{\mathsf{xy}} = 10$ ($\varphi_i = 1$)



(a) Pour l'image *Lena*

(b) Pour l'image *Paper*

(c) Pour l'image *Rose*

(d) Pour l'image *Baboon*

FIG. 1.4 – Impact de différentes attaques sur le tatouage avec information adjacente, en utilisant une énergie de marque du type $\sigma_{W_i} \propto \sigma_{X_i}$ telle que $D_{\mathsf{xy}} = 10$ ($\varphi_i = 1$)

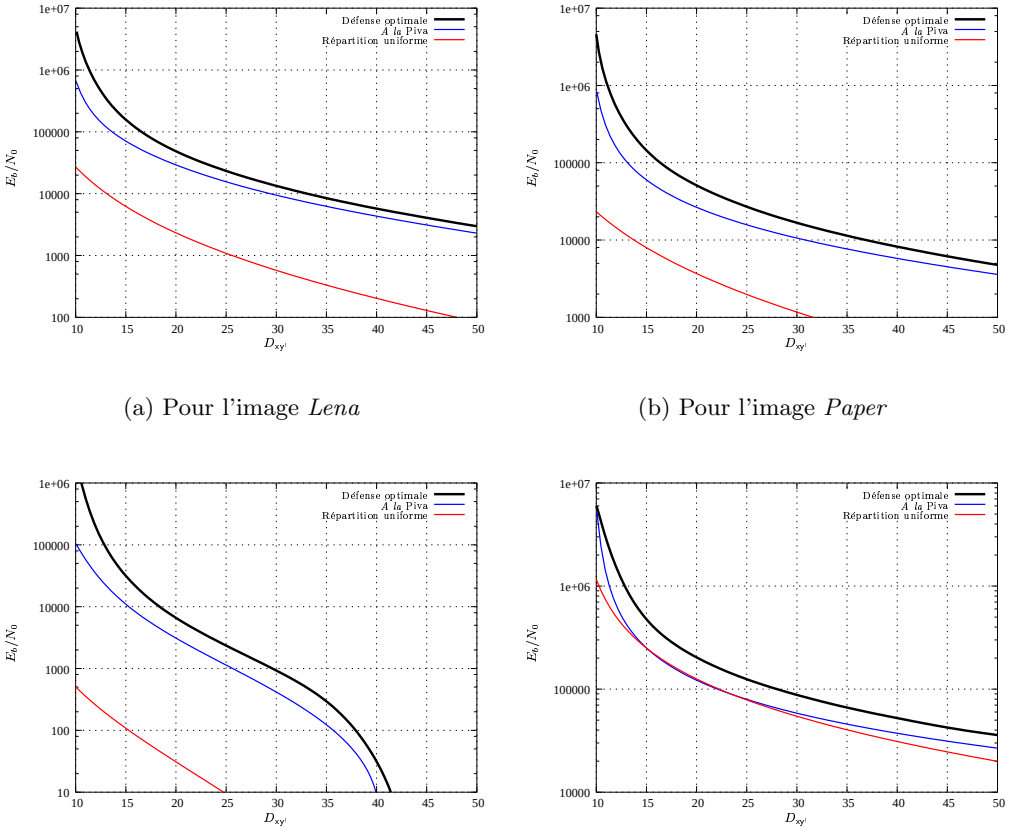
face à l'attaque optimale. Les paramètres de notre statégie d'insertion sont optimisés pour chaque niveau d'attaque $D_{\mathsf{xy}'}$. Comme prévu par l'optimisation utilisée, la défense optimale donne les meilleures performances pour les quatre images. De plus, face à des attaques autres (ajout de bruit gaussien uniforme ou proportionnel à l'énergie de la marque), les résultats sont encore meilleurs, comme le montre la figure 1.6.

Nous avons également utilisé le test Stirmark (voir la section 4.2.2 de la page 94 pour plus d'informations sur le mode opératoire) en fixant les paramètres λ et χ en vue d'une attaque de distorsion $D_{\mathsf{xy}'} = 40$. Les résultats sont présentés sur la figure 1.7. Comme pour les résultats de la partie précédente, on voit que cette famille d'attaque, pourtant largement utilisée dans la littérature pour démontrer la performance de techniques de tatouage, est sous-optimale et permet à notre stratégie d'insertion de fournir de bonnes performances.

Ces résultats montrent également l'intérêt de la prise en compte de l'information adjacente. Pour une distorsion d'attaque $D_{\mathsf{xy}'} = 40$, la capacité pour l'image *Lena* passe de 950 (résultats de la partie 2) à plus de 4000 bits (rapport $4,2$), tandis que pour *Baboon*, elle passe de 5000 à plus de 33000 bits (rapport $6,6$).

Conclusion

L'étalement de spectre est considéré comme la projection d'un signal sur un ensemble de porteuses, définissant ainsi un sous-espace linéaire. Les signaux ainsi projetés suivent une distribution i.i.d. et gaussienne, satisfaisant ainsi une des hypothèses nécessaires à l'application du schéma de Costa [Cos83]. La prise en compte de l'information adjacente permet de définir de nouvelles limites de capacité. Cela modifie considérablement la mesure de performance que nous utilisions dans la seconde partie de ce manuscrit. Les données du jeu entre l'attaquant et le défenseur sont donc modifiées. Nous avons résolu cette nouvelle optimisation et défini une attaque optimale et la défense associée en vue de l'utiliser dans un schéma prenant en compte l'information adjacente. Les résultats obtenus montrent que les stratégies ainsi définies apportent de forts gains par rapport aux autres stratégies testées.



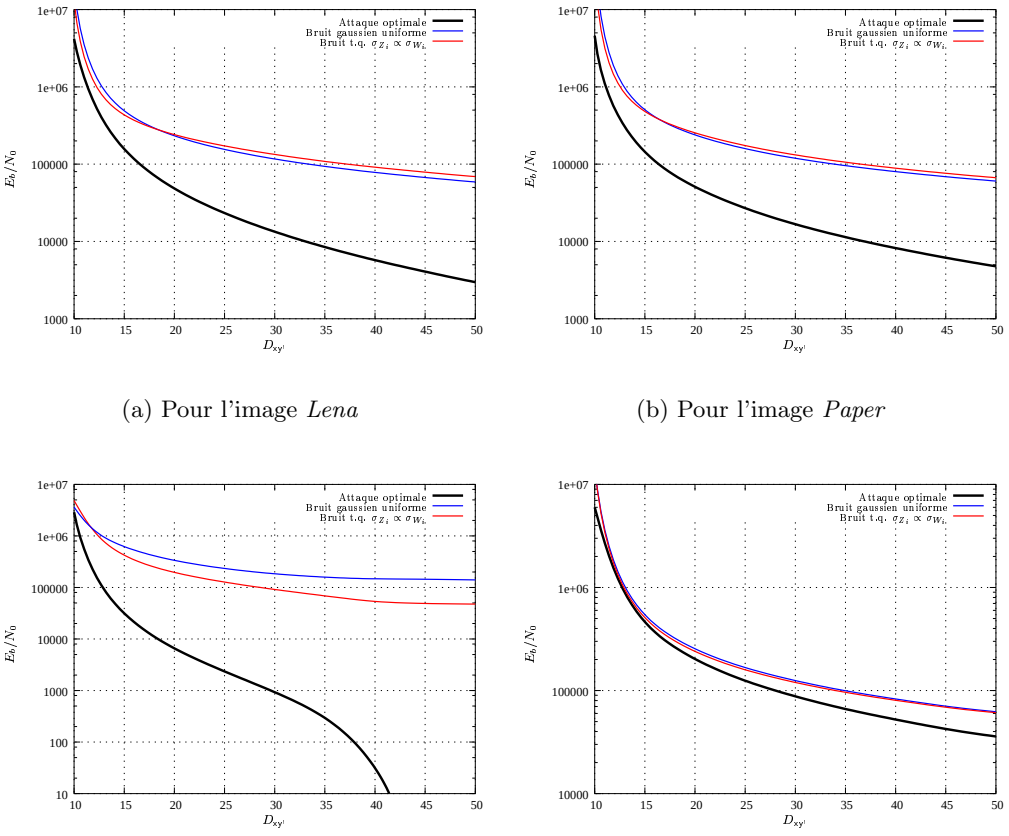
(a) Pour l'image *Lena*

(b) Pour l'image *Paper*

(c) Pour l'image *Rose*

(d) Pour l'image *Baboon*

FIG. 1.5 – Performances (avec prise en compte de l'information adjacente) de plusieurs répartitions de l'énergie de la marque face à l'attaque optimale. La distorsion d'insertion est fixée à $D_{\mathsf{xy}} = 10$ ($\varphi_i = 1$)



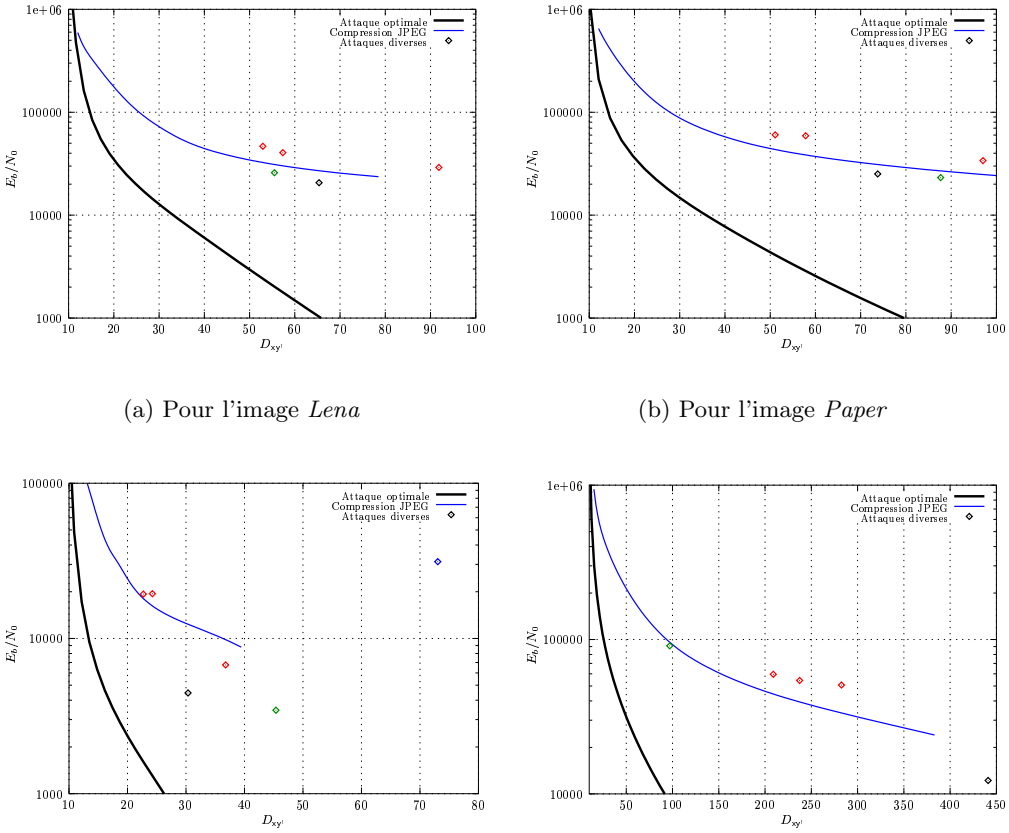
(a) Pour l'image *Lena*

(b) Pour l'image *Paper*

(c) Pour l'image *Rose*

(d) Pour l'image *Baboon*

FIG. 1.6 – Impact de l'ajout de bruit gaussien sur la performance de la défense optimale (avec information adjacente prise en compte). La distorsion d'insertion est fixée à $D_{\mathsf{xy}} = 10$ ($\varphi_i = 1$)



(a) Pour l'image *Lena*

(b) Pour l'image *Paper*

(c) Pour l'image *Rose*

(d) Pour l'image *Baboon*

FIG. 1.7 – Impact des attaques implémentées par Stirmark sur la performance de la défense optimale (avec information adjacente prise en compte). La distorsion d'insertion est fixée à $D_{\mathsf{xy}} = 10$ et nous visons une attaque de distorsion $D_{\mathsf{xy}'} = 40$ ($\varphi_i = 1$)

Chapitre 2

Codes pour le tatouage

La technique d'étalement de spectre que nous utilisons définit un sous-espace linéaire : le signal hôte de dimension m est projeté sur un ensemble de n porteuses. Dans cet espace, les signaux projetés suivent une loi Normale. Le signal marqué et le signal reçu sont respectivement donnés par

$$\mathsf{y}^{\mathrm{st}} = \mathsf{x}^{\mathrm{st}} + \mathsf{w}^{\mathrm{st}} \tag{2.1}$$
$$\mathsf{y}'^{\mathrm{st}} = \mathsf{x}^{\mathrm{st}} + \mathsf{w}^{\mathrm{st}} + \mathsf{z}^{\mathrm{st}}, \tag{2.2}$$

avec x^{st} et z^{st} modélisés par les v.a. $X^{\mathrm{st}} \sim \mathcal{N}(0, Q)$ et $Z^{\mathrm{st}} \sim \mathcal{N}(0, N)$. L'énergie de la marque est contrainte par $\sum_{j=1}^{n} \left[w_j^{\mathrm{st}}\right]^2 \leq nP$. Comme le signal projeté x^{st} est parfaitement connu lors de l'insertion, ce cas de figure correspond exactement à la définition d'un canal gaussien avec information adjacente disponible à l'encodeur. Les résultats vus dans la section 3.3 de la première partie peuvent donc être repris, y compris le schéma défini par Costa (l'ICS) permettant d'atteindre théoriquement la limite de capacité pour laquelle ont été optimisés les paramètres de l'étalement de spectre dans le chapitre précédent.

Appliquer directement le schéma de Costa est impossible pour des valeurs de n réalistes. Il utilise en effet un dictionnaire construit aléatoirement, et le décodage devrait donc se faire par recherche exhaustive. Nous allons proposer dans ce chapitre une technique permettant de construire le signal de marque w^{st} en s'appuyant sur un dictionnaire structuré construit à partir d'un système de codes correcteurs (ECC). La section 2.1 passe d'abord en revue quelques techniques représentatives s'appuyant sur des codes correcteurs. Puis nous présenterons dans la section suivante notre approche par codes poinçonnés. Enfin, la section 2.3 montrera une technique permettant de construire une marque w^{st} plus adaptée au tatouage que celle du schéma de Costa.

2.1 Dictionnaires structurés issus de codes correcteurs

La mise en œuvre pratique du schéma de Costa ne vient pas de la construction du dictionnaire structuré aléatoire, mais de son utilisation. Nous pouvons reprendre la

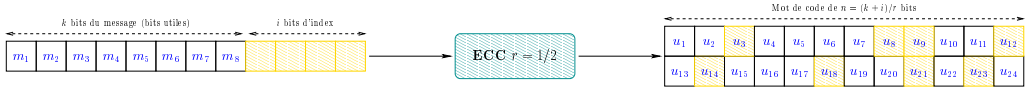


FIG. 2.1 – Ajout de i bits d'index afin de construire un dictionnaire structuré ($k = 8$, $i = 4$ et $n = 24$)

remarque faite sur les codes correcteurs dans la section 3.2 de la première partie concernant la complexité de décodage de signaux de dimensions importantes. Néanmoins, les systèmes de codes correcteurs utilisés aujourd'hui s'affranchissent de ce problème : codage par bloc pour limiter la taille de l'espace de recherche ou codage convolutif permettant d'obtenir un dictionnaire ordonné tout en conservant de bonnes performances. L'utilisation de codes correcteurs pourrait donc permettre la construction et l'exploitation d'un dictionnaire structuré tel que défini par Costa, tout comme ils permettent d'approcher les limites théoriques des canaux classiques. Dans la suite de cette section, nous nous plaçons dans le sous-espace défini par l'étalement de spectre. Les notations w^{st}, x^{st} et z^{st} sont respectivement simplifiées en w, x et z.

2.1.1 Approche par bits d'index et codes correcteurs

Le dictionnaire $\mathcal{U}$ utilisé dans le schéma de Costa est constitué de $2^{n(I(U;Y')-\epsilon)}$ mots de code[1]. Il est divisé en $2^{n(C-\epsilon)}$ sous-dictionnaires. Chaque message possible est donc associé à $2^{nI(U;X)}$ mots de codes. Une technique de construction simple pour obtenir un tel dictionnaire est d'utiliser des bits d'index afin de multiplier le nombre de mots de code par message. Nous considérons par la suite des messages binaires. Notons

$$i = n \times I(U;X) = \frac{n}{2}\log_2\left[1 + \frac{PQ}{(P+N)^2}\right]. \tag{2.3}$$

Si l'on insère ces i bits au sein du message que l'on souhaite transmettre et que l'on code le tout (par un code de rendement r), chaque message correspondra à 2^i mots de codes différents, ce qui satisfait les exigences de l'ICS. Cet enchaînement est montré par la figure 2.1.

La taille des mots de code ainsi générés sera de $(k+i)/r$. Or d'après l'équation 2.3, la valeur de i dépend du signal hôte x, et donc la taille des mots de code aussi. Mais en pratique la dimension n de x est fixe[2], et il est donc tout à fait possible que $n \neq (k+i)/r$: l'insertion $\mathsf{y} = \mathsf{x}+\mathsf{w}$ est impossible si les signaux à ajouter sont de dimensions différentes. Le rendement global k/n doit être fixe et indépendant de i (et donc de Q).

[1] Avec $\epsilon \to 0$ quand $n \to \infty$.

[2] Comme nous comptons utiliser ce type de dictionnaire au sein du sous-espace défini par étalement de spectre, nous pourrions faire en sorte que le nombre de porteuses n soit égal à la dimension des mots de codes. Or, en modifiant n, nous modifions Q et donc i. La dimension du sous-espace doit être adaptée jusqu'à convergence de i. Or elle n'est pas assurée dans tous les cas de figure.

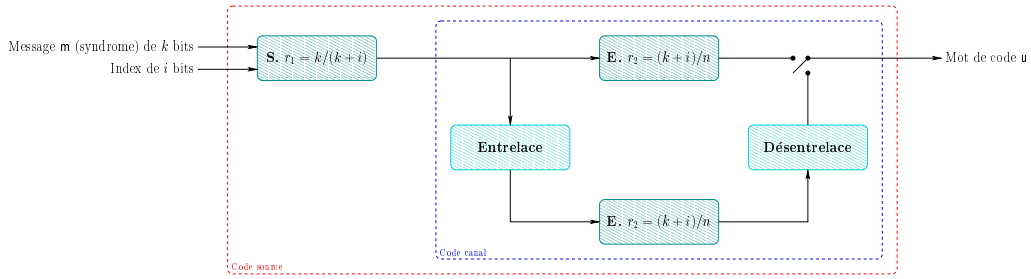


FIG. 2.2 – Construction du code structuré proposé par Chou *et al.* grâce à l'enchaînement d'un code par syndrome et de deux codes parallèles

2.1.2 Les codes de Chou *et al.*

Le codage par syndrome (voir la section 3.3.2 de la page 45) permet d'associer plusieurs mots de code à un même message. Il est de ce fait adapté à la construction d'un dictionnaire structuré pour le schéma de Costa. Chou et *al* [CPR99, CPR01] proposent d'utiliser le codage par syndrome associé à un code de type turbo (figure 2.2) afin d'assurer une bonne répartition des mots de codes au sein du dictionnaire global $\mathcal{U}$, mais aussi au sein des sous-dictionnaires $\mathcal{U}_{\mathsf{m}}$.

Un premier codeur utilise le principe des syndromes. Les i bits d'index sont encodés et les k bits du message forment le syndrome. D'un point du vue message[3], son rendement est $r_1 = k/(k+i)$: à un même message seront associés 2^i mots de code. Ce type de codeur assure une bonne répartition des mots de code d'un même sous-dictionnaire. Néanmoins, le pouvoir de correction est très faible car la distance minimale est de 1 (tous les éléments de $\{0,1\}^{k+i}$ sont des mots de code). Pour pallier ce défaut, un second étage de codage, utilisant deux codeurs en parrallèle, est appliqué. Son rendement est noté $r_2 = (k+i)/n$. Ce type de construction, associée à un décodage turbo, assure d'excellentes performances. Le rendement global est $r_1 \times r_2 = k/n$.

Toutefois, il est très difficile de faire varier i tout en gardant le rendement k/n constant. Les deux étages de codage doivent être adaptés, et il faut disposer d'une très large palette de codeurs pour assurer le rendement correct pour toute valeur de i (tableau 2.1).

2.2 Approche par codes poinçonnés

Comme nous l'avons déjà remarqué, les approches présentées ci-dessus ne peuvent garantir un rendement global constant malgré la variation probable de i suivant les signaux hôtes. Comme le nombre de mots de code par message possible varie en fonction de l'énergie de l'information adjacente à considérer, il est impossible d'assurer que la

[3]En fait, de par le principe du codage par syndrome, ce sont les i bits d'index qui sont encodés, le véritable rendement du code utilisé est donc $i/(k+i)$

i	r_1	r_2
0	1	1/3
1	64/65	65/192
2	32/33	33/96
3	64/67	67/192
4	16/17	17/48
...	...	...
32	2/3	1/2

TAB. 2.1 – Rendements des deux codeurs à utiliser pour obtenir un rendement k/n constant avec $k = 64$ bits et $n = 192$

FIG. 2.3 – Construction d'un motif pour l'encodage de $\mathsf{m} = \{10101010\}$ ($k = 8$ et $i = 4$)

dimension des mots de code du dictionnaire $\mathcal{U}$ soit la même que celle de x. Néanmoins, utiliser un système de codes correcteurs pour construire un dictionnaire adapté à l'ICS semble la solution la plus appropriée. Il faudrait donc un code dont le rendement puisse être augmenté ou diminué tout en gardant un bon pouvoir de correction. La solution se trouve du côté du poinçonnage. Le principe est tout simplement de supprimer une partie du mot de code afin d'adapter sa taille à celle désirée. Lors du décodage, les échantillons supprimés sont remplacés par des valeurs neutres afin de retrouver la taille originale du signal. Nous proposons d'utiliser ce principe au sein de codes convolutifs.

Considérons une valeur de i (obtenue par exemple en utilisant la formule 2.3), un message à transmettre m de k bits, un signal hôte de dimension n et un code de rendement r tel que $r = k/n$. L'encodage de m se fait en recherchant le mot de code $\mathsf{u}^\star$ le plus proche de x. Pour cela, nous construisons tout d'abord un motif comme illustré par la figure 2.3 : i bits d'index sont ajoutés aux bits du message et le tout est entrelacé afin de s'assurer d'une bonne répartition des mots de code associés à un même message. Ce motif va servir d'*a priori* fort dans un décodeur de Viterbi. À partir d'une matrice de poinçonnage $\mathsf{P} \in \mathcal{P}$ avec $\mathcal{P} = \{0,1\}^{(k+i)/r.n}$, nous définissons les fonctions

$$\begin{array}{rcl} \mathcal{X}^n \times \mathcal{P} & \longrightarrow & \mathcal{X}^{(k+i)/r} \\ \text{étend} : (\mathsf{x}, \mathsf{P}) & \longmapsto & \mathsf{x}^{\mathrm{e}} \end{array} \tag{2.4}$$

et

$$\begin{array}{rcl} \mathcal{X}^{(k+i)/r} \times \mathcal{P} & \longrightarrow & \mathcal{X}^n \\ \text{poinçonne} : (\mathsf{x}^{\mathrm{e}}, \mathsf{P}) & \longmapsto & \mathsf{x}. \end{array} \tag{2.5}$$

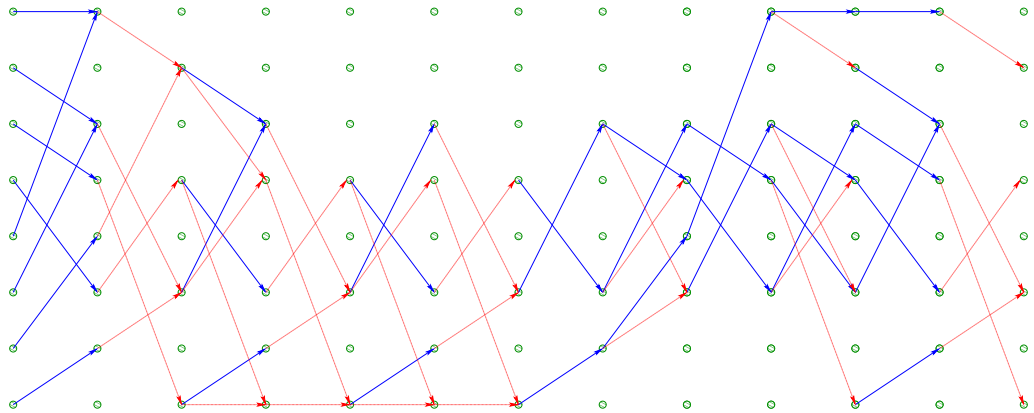


FIG. 2.4 – Treillis permettant de coder le message $\mathsf{m} = \{10101010\}$ en utilisant le motif de la figure précédente. Le treillis d'origine est donné par la figure 3.7(a) de la page 48

L'encodage du motif va aboutir à des mots de code de dimension $(k+i)/r$. Afin que la taille du signal hôte corresponde, ce dernier est étendu de i/r échantillons en utilisant des valeurs neutres pour le décodage (valeurs 0) grâce à l'utilisation de la fonction étend(). Il est ensuite décodé en utilisant le motif issu du message à transmettre (figure 2.5). Le décodage de x va donc fixer les i bits d'index de façon à ce que le mot de code obtenu soit le plus proche du signal x étendu et corresponde au message m. L'*a priori* va forcer certaines transitions du treillis (les k bits utiles du message), comme le montre la figure 2.4. Le mot de code issu du parcours du treillis est poinçonné avec poinçonne() afin de ramener sa dimension à n. C'est le mot de code $\mathsf{u}^\star$ et il est par construction le plus proche de x. À la réception des données y', le signal étend$(\mathsf{y}', \mathsf{P})$ est décodé en utilisant un algorithme de Viterbi classique (sans *a priori*). On identifie par la suite les k bits du message.

En utilisant une modulation de type BPSK[4] (le bit 0 devient -1 et le bit 1 devient $+1$), tous les mots de codes ainsi générés sont de même énergie. La performance de décodage est alors indépendante d'un éventuel facteur d'échelle de type $\mathsf{y}' = \gamma[\mathsf{x}+\mathsf{w}+\mathsf{z}]$.

Remarque

On voit dans cette approche une similitude avec le treillis de Miller *et al.* [MDC02], déjà vu dans la section 3.3.2 de la première partie. En regroupant les transitions libres (qui correspondent aux bits d'index du motif) avec les transitions $t-1$ (figure 2.6), la même propriété apparaît : pour un sommet et un bit donnés, plusieurs transitions sont possibles. On multiplie de ce fait le nombre de chemins par message.

[4]Pour *binary phase shift keying*.

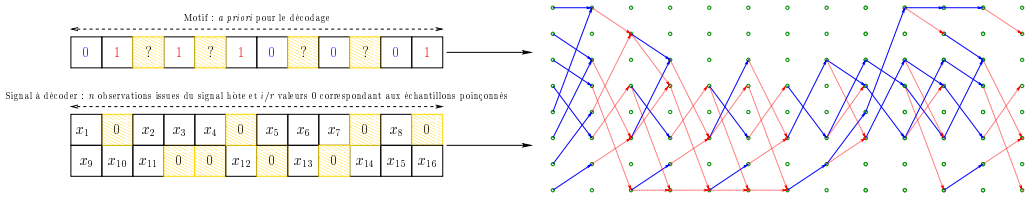


FIG. 2.5 – Encodage du message par décodage du signal hôte ($k = 8$, $i = 4$, $r = 1/2$ et $n = 16$)

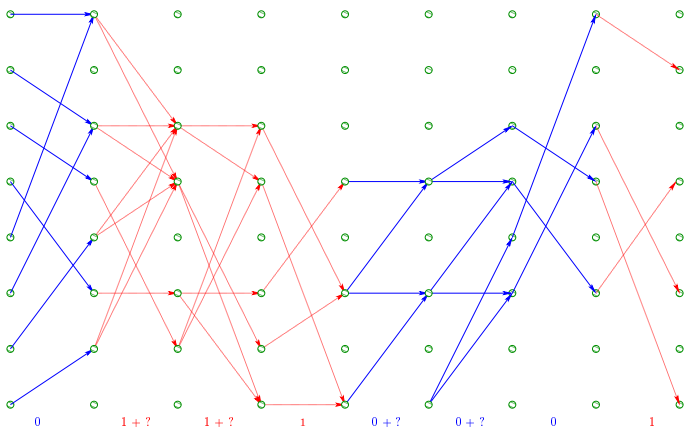


FIG. 2.6 – Treillis permettant de coder le message $\mathsf{m} = \{10101010\}$. Certains couples de transitions sont regroupés afin de montrer l'analogie avec le treillis de Miller

2.3 Maximisation de la robustesse

Costa indique dans son schéma la construction du signal à transmettre : $\mathsf{w} = \mathsf{u}^\star - \alpha\mathsf{x}$ avec $\alpha = P/(P+N)$. La valeur de α est celle qui maximise la capacité théorique du canal. Or, dans le cadre du tatouage robuste, cette notion de capacité reste abstraite : la taille du message est fixe et l'on souhaite maximiser la robustesse plutôt que maximiser la capacité. Nous allons d'abord montrer une interprétation géométrique de l'insertion d'une marque, puis nous verrons quelles techniques peuvent nous permettre de construire w avec les meilleures performances.

2.3.1 Interprétation géométrique

Les figures 2.8 et 2.9 donnent une interprétation de l'insertion d'une marque. Elles sont présentées en deux dimensions pour des raisons de lisibilité, mais doivent être considérées dans l'espace n-dimensionnel. Le dictionnaire est divisé en 2^{nC} sous-dictionnaires : dans l'exemple de la figure 2.8, il y a trois messages possibles, représentés par les couleurs bleu, rouge et vert. Chacun des mots de code (de même énergie) est associé à une zone de robustesse $\mathcal{R}_\mathsf{u}$ (qui a la forme d'un hyper-cône en n dimensions) à l'intérieur de laquelle le mot de code le plus proche est u. Ces zones sont délimitées par les traits pointillés de la figure. Supposons que l'on veuille transmettre le message bleu. Comme le défini le schéma de Costa, on recherche le mot de code $\mathsf{u}^\star$ le plus proche de x tel que $\mathsf{u}^\star \in \mathcal{U}_{\text{bleu}}$. Une fois cela fait, il faut construire un signal w tel que $\mathsf{x}+\mathsf{w}$ soit à l'intérieur de la zone de robustesse $\mathcal{R}_{\mathsf{u}^\star}$ et qu'il faille ajouter un bruit d'une énergie au moins égale à N pour en sortir, tout en respectant la contrainte d'énergie maximale (représentée par un cercle centré en x sur la figure).

La valeur de i permet d'adapter la répartition des mots de code. Si i est important, la taille des sous-dictionnaires sera importante. Et comme par construction ils sont supposés être répartis uniformément, la distance entre le signal hôte et le mot de code le plus proche sera statistiquement faible. D'un autre côté, la zone de robustesse associée sera peu étendue et l'énergie du bruit auquel il sera possible de résister (la robustesse) sera faible. À l'inverse, si l'on souhaite résister à un fort bruit, il vaut mieux augmenter la taille des zones de robustesse et donc rapprocher i de 0 (dictionnaire classique), comme le montre l'exemple de la figure 2.7.

2.3.2 Techniques d'insertion

L'ICS indique comment construire w de façon à optimiser la capacité du canal. Ses performances théoriques ne sont atteignables que lorsque le nombre d'échantillons n tend vers l'infini. Or, en pratique, ce nombre est bien sûr limité. Les paramètres d'insertion α et i donnés par Costa ne sont peut-être pas optimaux pour des cas réalistes. De plus, nous ne recherchons pas à maximiser la capacité du canal mais la robustesse de la transmission (notre capacité est fixée et nous voulons maximiser N, alors que Costa maximise la capacité pour un N donné). Nous devons donc adapter la construction de w au code dont nous disposons et à notre recherche de robustesse.

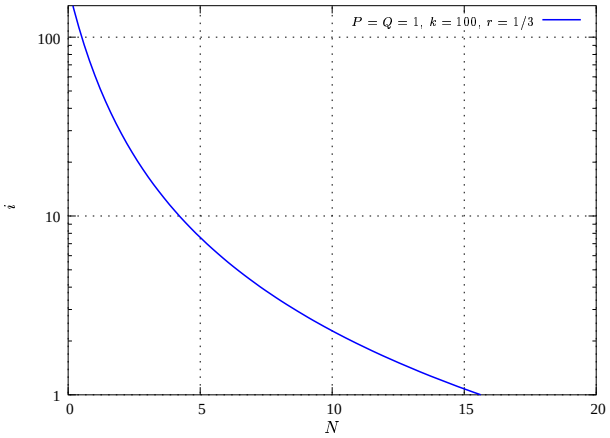


FIG. 2.7 – Nombre de bits d'index à ajouter (équation 2.3) en fonction du bruit d'attaque

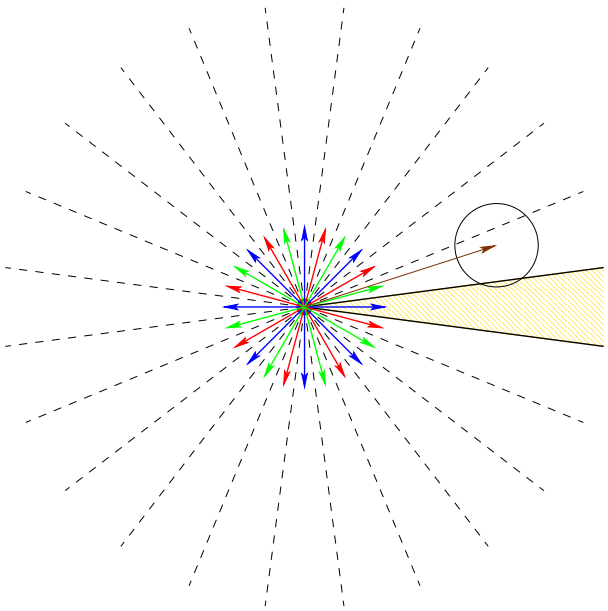


FIG. 2.8 – Répartition des mots de codes dans l'espace et zones de robustesse

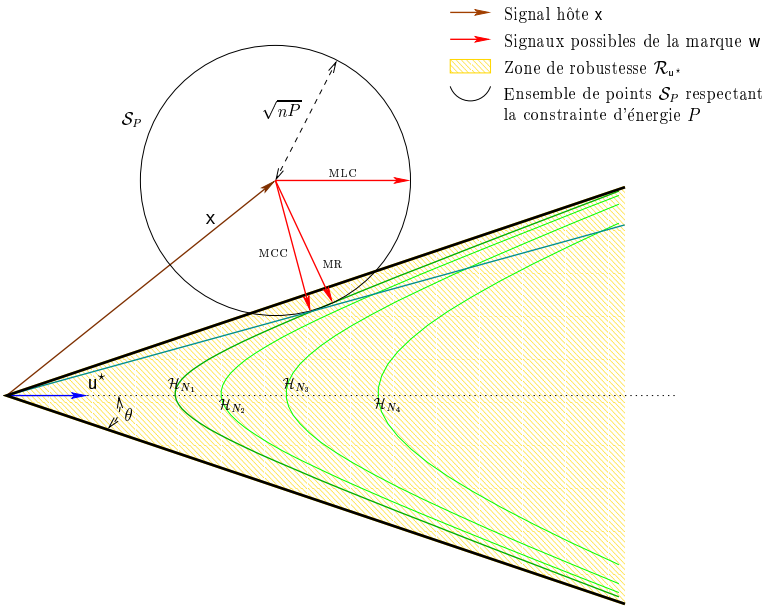


FIG. 2.9 – Différentes techniques de construction du signal de la marque w. Figure inspirée de celle présentée par Miller *et al.* [MCB00]

La figure 2.9 montre plus précisement la phase d'insertion. La contrainte de puissance sur w défini une hyper-sphère $\mathcal{S}_P$ de rayon $\sqrt{nP}$ centrée en x. Tous les points de $\mathcal{S}_P$ sont des candidats potentiels pour y. Or, il faut bien sûr que y appartienne à $\mathcal{R}_{u^\star}$ afin d'assurer un bon décodage. La recherche peut donc se limiter à $\mathcal{R}_{u^\star} \cap \mathcal{S}_P$.

Nous pouvons mesurer la robustesse d'un point par la distance statistique de ce point par rapport aux bords de l'hyper-cône. L'ensemble des points situés à une distance N (il faut ajouter un bruit d'énergie N pour quitter l'hyper-cône) définit un hyperboloïde, dont plusieurs exemples sont donnés sur la figure 2.9 :

$$\mathcal{H}_N = \left\{ y \text{ t.q. } N = \left[\frac{\langle y, u^\star \rangle}{\|u^\star\|} \right]^2 (1 + \tan^2 \theta) - \|y\|^2 \right\}, \tag{2.6}$$

où θ est l'angle de l'hyper-cône $\mathcal{R}_{u^\star}$. Nous pouvons déduire de la théorie du *sphere packing* [Lee67] la formule

$$\tan^{-2} \theta = 2^{\frac{2(k+i)}{n}} - 1. \tag{2.7}$$

La marque $w = u^\star - Px/(P+N)$ que définit Costa est optimale dans le pire cas[5] (c'est-à-dire que x est situé à équidistance de deux mots de codes du sous-dictionnaire du message à transmettre). Dans d'autres cas, il se peut que la marque de Costa ne soit pas optimale en terme de robustesse et qu'il soit possible de proposer mieux.

[5]Une justification géométrique de ces paramètres est donnée en annexe B.

Miller *et al.* [MCB00] ont énuméré trois[6] principales techniques d'insertion en vue d'une utilisation dans le cadre de la détection de marque (ces techniques sont illustrées par les flèches rouges de la figure 2.9). Les auteurs se basent sur une zone de détection (si le signal reçu est dans cette zone, la marque est considérée comme présente) que l'on peut assimiler dans notre cas à la zone de robustesse $\mathcal{R}_{\mathsf{u}^\star}$:

- la première méthode consiste à prendre $\mathsf{w} \propto \mathsf{u}^\star$, et correspond en fait à une transmission sans prise en compte de l'information adjacente (insertion aveugle). Or, comme le montre la figure 2.9, il est possible que y ne soit pas dans la zone de robustesse et que le message extrait soit faux même en l'absence de bruit d'attaque z. Miller baptise cette méthode MLC (*maximizing linear correlation*). D'après la formulation de Costa, cela équivaut à faire tendre α vers zéro. Cette technique maximise la mesure $\langle \mathsf{u}^\star, \mathsf{y} \rangle$,
- la technique MCC (*maximizing correlation coefficient*) consiste à se rapprocher le plus possible du mot de code. S'il y a intersection entre l'hyper-cercle définissant la limite de distorsion et $\mathsf{u}^\star$ (par exemple si $P > Q$), cela équivaut à $\mathsf{y} \propto \mathsf{u}^\star$. Le signal hôte est alors totalement supprimé[7]. La mesure

$$c = \frac{\langle \mathsf{u}^\star, \mathsf{y} \rangle}{\|\mathsf{u}^\star\| \times \|\mathsf{y}\|}$$

  est maximisée,
- la dernière technique (notée MR pour *maximum robustness*) consiste à trouver le signal w qui maximise l'énergie N du bruit qu'il est possible de supporter, c'est-à-dire à trouver le point tangent à $\mathcal{H}_N$ tel que N soit le plus fort possible. Sur l'exemple de la figure 2.9, cette technique permet de résister à l'ajout d'un bruit d'énergie N_1.

Faire en sorte que le signal marqué soit le plus loin possible des limites de la zone de robustesse est la technique qui a le plus de sens dans le cadre du tatouage robuste. En l'absence de connaissance sur le bruit d'attaque ajouté, et d'après l'équation 2.6, on peut définir w par

$$\mathsf{w} = \arg \max_{\tilde{\mathsf{w}} \text{ t.q. } \|\tilde{\mathsf{w}}\|^2 = nP} \left\{ \left[\frac{\langle \mathsf{x} + \tilde{\mathsf{w}}, \mathsf{u}^\star \rangle}{\|\mathsf{u}^\star\|} \right]^2 \left(1 + \tan^2 \theta\right) - \|\mathsf{x} + \tilde{\mathsf{w}}\|^2 \right\}. \tag{2.8}$$

Pour une valeur i donnée, la maximisation se fait en recherchant sur la sphère $\mathcal{S}_P$ le point $\mathsf{y} = \mathsf{x} + \mathsf{w}$ maximisant N.

Remarque

Le schéma de Costa nous indique la taille optimale des sous-dictionnaires. Nous en déduisons le nombre de bits d'index i à utiliser pour notre code structuré (équation 2.3).

[6] Dans cet article, une quatrième technique est présentée, notée CR pour *constant robustness*. Elle cherche à garantir une robustesse constante sans se soucier de la contrainte de distorsion, et ne respecte donc pas nos hypothèses.

[7] Cette technique est également connue sous le nom de *precancellation* [PS98, CW00, CMB02].

Reste que ce paramètre doit être connu lors du décodage (cela est également vrai pour les codes proposés par Chou *et al.*) et doit donc être transmis. On retrouve la même problématique que le tatouage par quantification : il faut transmettre les paramètres permettant de construire le dictionnaire utilisé à l'insertion. Plusieurs solutions sont envisageables, comme la réservation des premiers symboles de w^{st}, ou alors l'ajout d'une marque supplémentaire pour transmettre i.

2.4 Suppression de l'interférence inter-symboles

L'optimisation du chapitre 1 se base sur l'hypothèse que le bruit du signal hôte n'intervient pas dans la mesure de performance, grâce à l'utilisation de codes adaptés utilisant l'information adjacente disponible. Or la projection par étalement de spectre introduit un bruit supplémentaire. Reprenons l'équation de la projection des données reçues (marquées et attaquées) :

$$\begin{aligned} y'^{\mathrm{st}}_j &= \sum_{i=1}^{m} \beta_i \times \mathsf{G}(i,j) \times y'_i \\ &= \sum_{i=1}^{m} \beta_i \times \mathsf{G}(i,j) \left[\gamma_i \left[x_i + \frac{\sigma_{W_i}}{\|\mathsf{w}^{\mathrm{st}}\|} \sum_{k=1}^{n} \mathsf{G}(i,k) \times w^{\mathrm{st}}_k \right] + z_i \right] \\ &= x^{\mathrm{st}}_j + w^{\mathrm{st}}_j + z^{\mathrm{st}}_j \\ &\quad + \sum_{i=1}^{m} \left[\mathsf{G}(i,j) \frac{\beta_i \gamma_i \times \sigma_{W_i}}{\|\mathsf{w}^{\mathrm{st}}\|} \sum_{k=1, k\neq j}^{n} \mathsf{G}(i,k) \times w^{\mathrm{st}}_k \right]. \end{aligned} \tag{2.9}$$

Le dernier terme de la somme est l'interférence des autres éléments de w^{st} et est connu sous le nom d'interférence inter-symboles (ISI). Si les porteuses sont générées pseudo-aléatoirement, c'est un bruit i.i.d. suivant une loi Normale de moyenne nulle et dont la variance est

$$I = \sum_{i=1}^{m} \beta_i^2 \gamma_i^2 \times \sigma^2_{W_i} \frac{n-1}{n}. \tag{2.10}$$

Pour l'étalement de spectre classique vu dans la seconde partie de ce manuscrit, sans prise en compte de l'information adjacente, cette interférence a peu d'influence sur la capacité du canal. On passe d'un rapport $P/(N+Q)$ au rapport $P/(N+Q+I)$ avec $I \ll Q$. Comme le montre la figure 2.10 (courbes rouges), l'impact de l'interférence inter-symboles sur la performance est alors très faible et quasi négligeable. Par contre, s'il y a prise en compte du signal hôte à l'insertion, on passe d'un rapport P/N à un rapport $P/(N+I)$: la perte est importante (courbes bleues de la figure 2.10). Les performances présentées à la fin du chapitre 1 ne prennent pas en compte l'ISI. Il faut donc un moyen de la supprimer pour espérer les atteindre.

2.4.1 Porteuses orthogonales

L'interférence est due au fait que les porteuses ne sont pas parfaitement orthogonales. Un moyen simple de l'éviter est de transmettre un seul élément de w^{st} par échantillon [CW01], c'est-à-dire que $\forall i \in \{1, 2, \ldots, m\}$, il faut qu'il y ait un seul élément non nul dans $\{\mathsf{G}(i,1), \mathsf{G}(i,2), \ldots, \mathsf{G}(i,n)\}$. Les porteuses sont alors orthogonales et $I = 0$. Néanmoins, cette technique limite l'étalement : chaque élément de w^{st} est étalé sur m/n échantillons (et non sur m), ce qui pose problème si n est important. De plus, pour un signal hôte de faible énergie (par exemple l'image *Rose* que nous utilisons pour les expérimentations présentées ici), le nombre d'échantillons pouvant accueillir une marque de forte énergie est limité car la marque est principalement située sur les coefficients aux variances les plus fortes (voir la forme de l'insertion présentée par la figure 3.1 de la page 87). De ce fait, il y a risque que l'énergie globale ne soit pas équitablement partagée par les éléments de w^{st}. Les signaux projetés ne seront probablement pas i.i.d. : l'utilisation de canaux parallèles serait plus appropriée.

2.4.2 L'interférence inter-symboles comme information adjacente

Nous proposons de prendre en compte l'interférence inter-symboles lors du codage du message et de la construction de w^{st} afin de l'ajouter à l'information adjacente. Nous devons construire un signal w^{st} prenant en compte le signal hôte x^{st} et sa propre interférence. Cela correspond à la résolution d'un système linéaire de dimension $n \times n$. Nous proposons d'utiliser une résolution itérative de type Jacobi, présentée par l'algorithme 2.1. Comme l'ISI est faible par rapport au signal hôte, nous pouvons initialiser l'algorithme en approximant la solution sans prendre en compte l'ISI. Nous en déduisons le signal w. Son interférence est ajoutée au signal hôte : $\mathsf{x}^{\mathrm{st}} \leftarrow \mathsf{x}^{\mathrm{st}} + \mathrm{isi}(\mathsf{w}^{\mathrm{st}})$. Cette information adjacente est utilisée pour raffiner w^{st}. Le processus est répété jusqu'à ce que la valeur de w^{st} se stabilise. En pratique, la convergence est obtenue en moins de trois itérations.

2.5 Résultats

Cette section présente les résultats obtenus par application de notre dictionnaire structuré dans le cadre d'un canal avec information adjacente. Les deux premières expérimentations utilisent une chaîne de transmission comme celle décrite par la figure 2.11 : le message m est codé en utilisant le dictionnaire défini dans la section 2.2, et un signal de marque w respectant la contrainte d'énergie P est transmis. Il est bruité par x et z. Le signal reçu y' est décodé. La dernière expérience montre l'intérêt de la prise en compte de l'interférence inter-symboles dans le cadre du tatouage avec information adjacente.

2.5.1 Comparaison entre codes structurés et codes classiques

Nous comparons tout d'abord notre approche à un code classique non-structuré. Nous disposons de deux codes correcteurs convolutifs binaires : un code de rendement

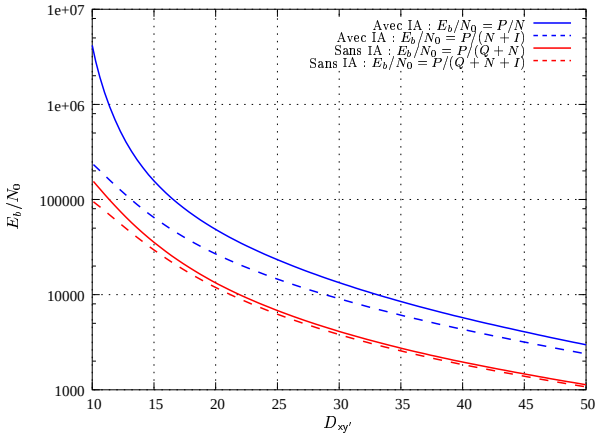


FIG. 2.10 – Influence de l'interférence inter-symboles sur le tatouage basé sur l'étalement de spectre (image *Lena* de taille 512×512, $n = 100$, $\varphi_i = 1$ et $D_{\mathsf{xy}} = 10$)

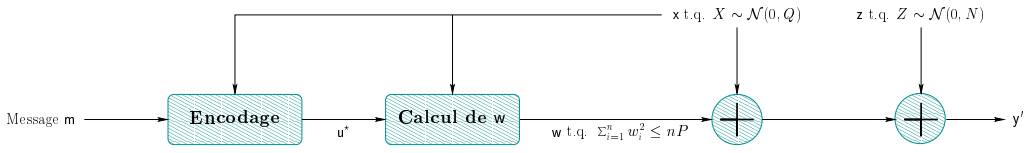


FIG. 2.11 – Transmission du message m sur un canal gaussien avec l'information x disponible à l'encodage

for $j = 1$ to n **do**
$x_j^{\mathrm{st}} \leftarrow \sum_{i=1}^{m} \beta_i \gamma_i \times \mathsf{G}(i,j) \times x_i$
end for

$\mathsf{u}^\star \leftarrow$ mot de code le plus proche de x^{st}
$\widetilde{\mathsf{w}}^{\mathrm{st}} \leftarrow \arg \max_{\mathsf{w} \text{ t.q. } \|\mathsf{w}\|^2 = nP} \left\{ \left[\frac{\langle \mathsf{x}+\mathsf{w}, \mathsf{u}^\star \rangle}{\|\mathsf{u}^\star\|} \right]^2 \left(1+\tan^2\theta\right) - \|\mathsf{x}+\mathsf{w}\|^2 \right\}$
repeat
$\mathsf{w}^{\mathrm{st}} \leftarrow \widetilde{\mathsf{w}}^{\mathrm{st}}$

for $j = 1$ to n **do**
$x_j^{\mathrm{st}} \leftarrow 0$
for $i = 1$ to m **do**
$w_i \leftarrow \frac{\sigma_{W_i}}{\|\mathsf{w}^{\mathrm{st}}\|} \sum_{k=1}^{n} \mathsf{G}(i,k) \times w_k^{\mathrm{st}}$
$\mathsf{I}(i,j) \leftarrow w_i - \frac{\sigma_{W_i}}{\|\mathsf{w}^{\mathrm{st}}\|} \mathsf{G}(i,j) \times \widetilde{w}_j^{\mathrm{st}}$
$x_j^{\mathrm{st}} \leftarrow x_j^{\mathrm{st}} + \beta_i \gamma_i \left(x_i + \mathsf{I}(i,j)\right) \mathsf{G}(i,j)$
end for
end for

$\mathsf{u}^\star \leftarrow$ mot de code le plus proche de x^{st}
$\widetilde{\mathsf{w}}^{\mathrm{st}} \leftarrow \arg \max_{\mathsf{w} \text{ t.q. } \|\mathsf{w}\|^2 = nP} \left\{ \left[\frac{\langle \mathsf{x}+\mathsf{w}, \mathsf{u}^\star \rangle}{\|\mathsf{u}^\star\|} \right]^2 \left(1+\tan^2\theta\right) - \|\mathsf{x}+\mathsf{w}\|^2 \right\}$
until $|\widetilde{\mathsf{w}}^{\mathrm{st}} - \mathsf{w}^{\mathrm{st}}| \leq \epsilon$

Alg 2.1: Calcul de w^{st} afin de prendre en compte l'interférence inter-symboles en plus de l'information adjacente et de maximiser la robustesse

1/2 et un autre de rendement 1/3. Ils ont tous les deux une longueur de contrainte (taille du registre à décalage) de 9. Nous utilisons les paramètres de Costa, c'est-à-dire $\alpha = P/(P+N)$ et i défini selon l'équation 2.3. Cette équation montre que si $Q \geq P\left[4^{1-r}-1\right]$ alors le nombre de bits d'index est supérieur au nombre de bits de redondance pour les plus petites valeurs de N. Il est donc impossible de transmettre correctement les bits utiles. Pour cette raison, nous avons limité l'énergie de x à $Q = 0,75 \times P$ pour $r = 1/2$ et à $Q = 1,15 \times P$ pour $r = 1/3$. Cela n'est pas génant dans notre cadre d'utilisation. En effet, l'étalement de spectre fait que l'énergie de la marque est concentrée dans le sous-espace. Ainsi, les valeurs de Q restent limitées par rapport à celles de P. Par exemple, en prenant $D_{\mathsf{xy}} = 10$ et $n = 100$ et en utilisant les formules 1.8 et 1.10 (page 107), nous avons $Q/P \simeq 1,1 \times 10^{-2}$ pour *Lena*, environ $2,5 \times 10^{-2}$ pour *Paper* et *Baboon* et $6,8 \times 10^{-3}$ pour *Rose*. L'énergie du signal transmis w est fixée à $P = 1$.

La figure 2.12 montre les taux d'erreur par bit obtenus. Le rapport signal-à-bruit du canal normalisé par le rendement (afin de pouvoir comparer les deux codes de rendements différents) est noté en abscisse et la probabilité d'erreur en ordonné. Pour les trois valeurs de Q testées, nous avons ajouté à titre de comparaison les performances du code de base soumis à la somme des bruits $\mathsf{x} + \mathsf{z}$ (traits pointillés). Nous avons de plus indiqué (courbe en trait plein épais) la performance du code classique sur un canal perturbé uniquement par z : cette courbe mesure la performance maximale théorique que pourrait atteindre le code (suppression parfaite de l'influence de x sur les performances).

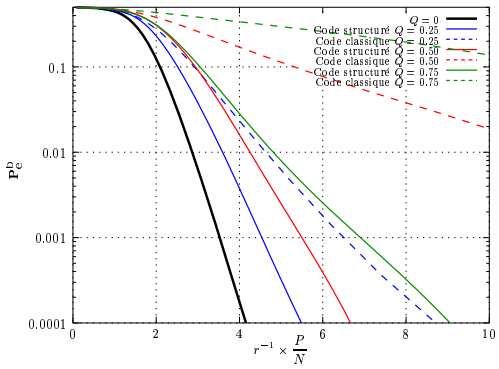
L'utilisation de codes structurés associée à la prise en compte de l'information adjacente apporte un net gain de performance (par exemple, gain de 2 dB à $\mathbf{P}_{\mathrm{e}}^{\mathrm{b}} = 10^{-4}$ sur le code de rendement 1/2 pour $Q = 0,25$). L'influence de x sur $\mathbf{P}_{\mathrm{e}}^{\mathrm{b}}$ est réduite. Néanmoins, la robustesse reste dépendante de Q : plus l'énergie de l'information adjacente est élevée et plus la performance est éloignée de l'idéal théorique (pour $r = 1/2$, on passe d'environ 1,2 dB de différence entre l'idéal et les performances pratiques pour $Q = 0,25$ à une différence de 2 dB pour $Q = 0,5$). La figure 2.13 reprend la même expérimentation mais indique les probabilités d'erreur par message.

2.5.2 Techniques d'insertion

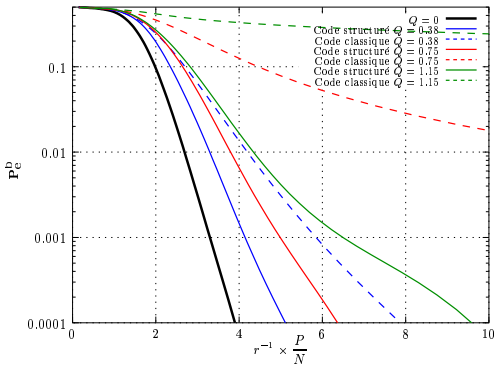
Nous testons ensuite les différentes techniques d'insertion vues dans la section 2.3.2. Nous utilisons notre dictionnaire structuré en nous appuyant sur des codes convolutifs de rendements 1/2 ou 1/3 comme pour les tests précédents. Le signal w transmis est construit avec

- la technique MLC,
- la technique MCC,
- la méthode donnée par Costa : $\alpha = P/(P+N)$,
- la technique MR.

Son énergie est fixée à $P = 1$. La valeur i est calculée par l'équation 2.3.

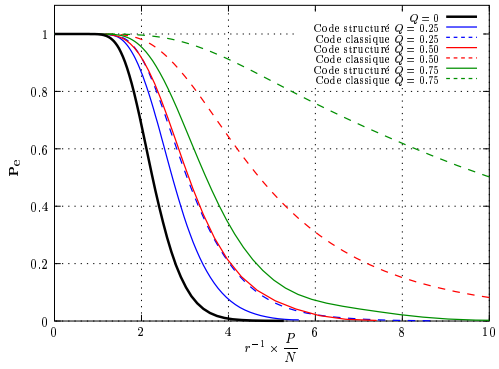


(a) Code de rendement $r = 1/2$ et de longueur de contrainte 9

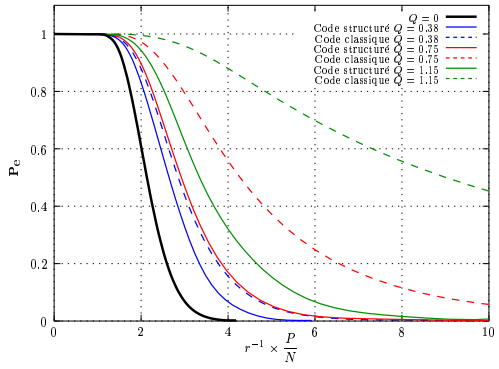


(b) Code de rendement $r = 1/3$ et de longueur de contrainte 9

FIG. 2.12 – Probabilité d'erreurs par bit en utilisant des codes convolutifs structurés ou non

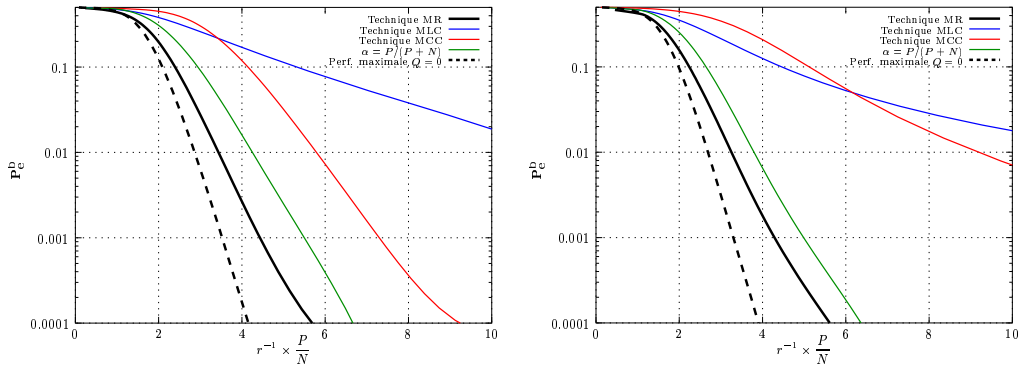


(a) Code de rendement $r = 1/2$ et de longueur de contrainte 9



(b) Code de rendement $r = 1/3$ et de longueur de contrainte 9

FIG. 2.13 – Probabilités d'erreur par message en utilisant des codes convolutifs structurés ou non ($k = 100$ bits)



(a) Code de rendement $r = 1/2$ et de longueur de contrainte 9, avec une information adjacente telle que $Q = 0.5$

(b) Code de rendement $r = 1/3$ et de longueur de contrainte 9, avec une information adjacente telle que $Q = 0.75$

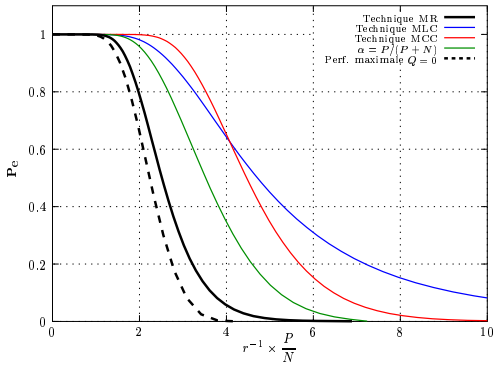
FIG. 2.14 – Probabilités d'erreur par bit pour différentes méthodes de construction de w

La figure 2.14 montre les résultats obtenus. L'abscisse des graphiques indique le rapport signal-à-bruit normalisé par le rendement et l'ordonné la probabilité d'erreur par bit. En plus des quatre techniques listées ci-dessus, nous avons ajouté la performance maximale théorique du code (voir l'explication dans la section précédente). On voit que la technique de Costa est sous-optimale dans notre cas et qu'il est plus intéressant de maximiser la robustesse. La différence entre les résultats pratiques et l'idéal théorique passe de 2 à 1,2 dB pour le code de rendement 1/2 à $\mathbf{P}_e^b = 10^{-4}$. L'analyse de la figure 2.15 délivre les mêmes conclusions en ce qui concerne la probabilité d'erreur par message.

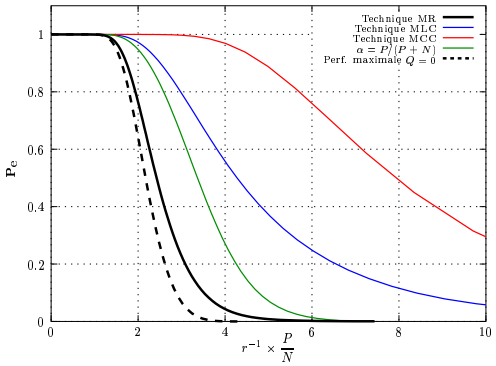
2.5.3 Gains grâce à la prise en compte de l'ISI

Nous reprenons le mode opératoire des expériences du chapitre 1 de cette troisième partie. Nous utilisons la stratégie de défense optimale et appliquons trois attaques : attaque optimale, ajout de bruit gaussien uniforme et de bruit gaussien proportionnel à l'énergie de la marque.

La figure 2.16 montre la différence de performance entre tatouage avec prise en compte complète de l'information adjacente (signal hôte et ISI) et tatouage avec prise en compte uniquement du signal hôte. Les gains les plus importants sont obtenus pour les attaques les plus faibles (pour l'image *Baboon* et contre l'attaque optimale de distorsion $D_{\mathsf{xy}'} = 15$ (PSNR de 36,4 dB), la prise en compte de l'ISI fait passer d'une capacité maximale de 85000 bits à plus de 175000 bits, soit un rapport 2). La différence s'amenuise pour les distorsions importantes, car le bruit d'attaque devient plus important que le bruit de l'ISI (à $D_{\mathsf{xy}'} = 45$, on passe de 22500 à 28000 bits, soit un rapport 1,25).



(a) Code de rendement $r = 1/2$ et de longueur de contrainte 9, avec une information adjacente telle que $Q = 0.5$



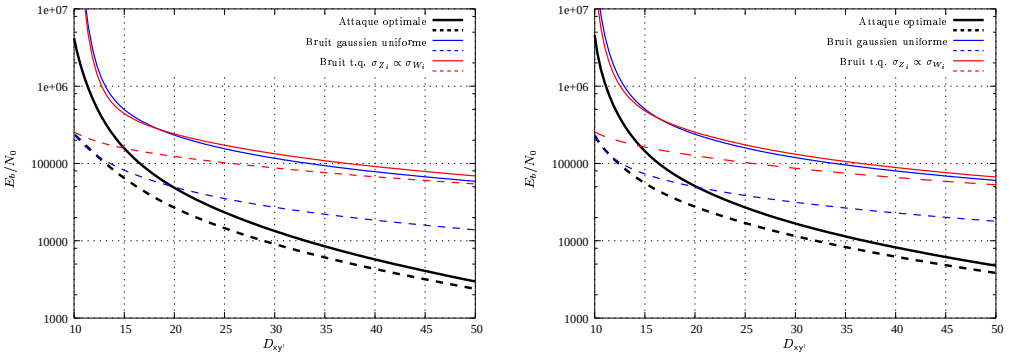
(b) Code de rendement $r = 1/3$ et de longueur de contrainte 9, avec une information adjacente telle que $Q = 0.75$

FIG. 2.15 – Probabilités d'erreur par message pour différentes méthodes de construction de w ($k = 100$ bits)

Conclusion

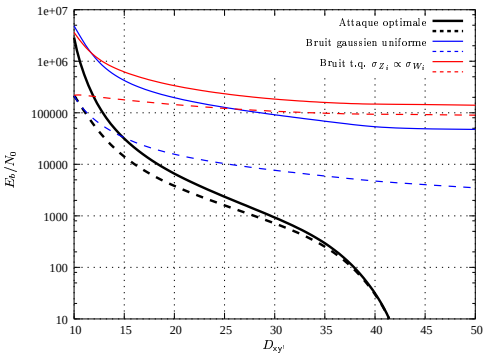
La limite de capacité de Costa est démontrée en utilisant un dictionnaire structuré aléatoire et en considérant une dimension $n \to \infty$. On retrouve ce type de construction dans les calculs de capacité de canaux classiques. L'utilisation de codes correcteurs en vue de construire un dictionnaire structuré semble être logique. Néanmoins, le rendement global du code doit être constant, malgré l'addition de bits servant à indexer les différents mots de code par message : il nous faut donc un code correcteur de rendement variable.

La solution est apportée par les codes poinçonnés. À partir d'un code convolutif classique, nous avons développé une technique permettant de multiplier le nombre de mots de code par message, de rechercher le mot de code $\mathsf{u}^\star$ le plus proche d'un signal hôte donné et de décoder le signal reçu, tel que l'a défini Costa. Les résultats nous montrent que dans le cas de canaux avec information adjacente, ce code est bien plus performant qu'un code classique équivalent, malgré la perte de pouvoir correcteur due au poinçonnage. De plus, nous avons montré que les paramètres de Costa (forme du signal à transmettre) ne sont pas les plus adaptés au problème du tatouage robuste. Il vaut mieux maximiser la robustesse au cas par cas lors de l'insertion. Les expérimentations montrent que les performances atteintes dans ce cas de figure approchent l'idéal théorique. Enfin, nous avons exposé un algorithme permettant de supprimer l'interférence inter-symboles à la construction du signal de la marque et d'obtenir des gains significatifs en terme de performance.

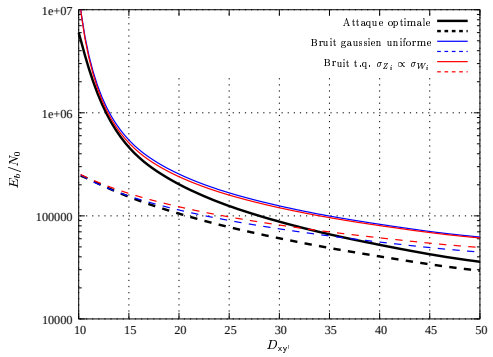


(a) Pour l'image *Lena*

(b) Pour l'image *Paper*



(c) Pour l'image *Rose*



(d) Pour l'image *Baboon*

FIG. 2.16 – Gains apporté par la prise en compte de l'interférence inter-symboles (courbes en traits pleins contre courbes en pointillés). La stratégie d'insertion est celle définie par max-min dans le chapitre précédent et la distorsion d'insertion est fixée à $D_{\mathsf{xy}} = 10$ ($\varphi_i = 1$)

Chapitre 3

Raffinements du jeu

Le premier chapitre de cette partie a exposé une stratégie d'insertion adaptée aux canaux avec prise en compte de l'information adjacente. Elle a été obtenue par application de la théorie des jeux, grâce à une optimisation de type max-min. Nous proposons maintenant d'affiner ce jeu en utilisant des mesures différentes.

La première section introduit dans la mesure de performance E_b/N_0 l'influence d'une erreur de recalage. Le jeu est alors modifié et les stratégies résultantes sont appliquées aux signaux issus d'une transformée en ondelettes. La seconde section utilise des mesures de distorsion prenant en compte la réalisation du signal. Nous avions pour le moment utilisé des mesures basées sur la distribution statistique du signal hôte (variances de chacun des échantillons de x), connue par les deux protagonistes du jeu. Or l'attaquant connaît en plus le signal y et le défenseur le signal x.

3.1 Introduction de la désynchronisation géométrique

Les attaques géométriques consistent à désynchroniser l'extracteur. Par une transformation géométrique (rotations, translations, déformations locales telles que celles implémentées par Stirmark [Pet00, sti], ...), le signal après attaque y' est décalé. Si une extraction comme celle présentée ici (basée sur la corrélation entre les porteuses de G et le signal y') est faite, il y a peu de chances de retrouver le message correct. Un schéma de tatouage dont une des priorité est la résistance aux attaques géométriques doit inclure un module de recalage, comme celui illustré par la figure 3.1.

Les techniques de recalage proposées dans la littérature se séparent en deux principales catégories. La première approche est d'utiliser un domaine invariant aux attaques géométriques visées [RP98, PRD+99], comme la transformée de Fourier-Mellin déjà évoquée dans la section 2.1.1 de la première partie. La seconde possibilité est d'introduire au sein du signal marqué un motif de synchronisation [PRD+99, VDP01] aux propriétés géométriques connues (pics d'auto-corrélation, structures périodiques, ...). Il peut être directement introduit dans la marque w ou grâce à un signal supplémentaire. En identifiant les caractéristiques du motif depuis l'image attaquée, la transformation géométrique subie peut être estimée et inversée. Néanmoins, dans les deux cas, les

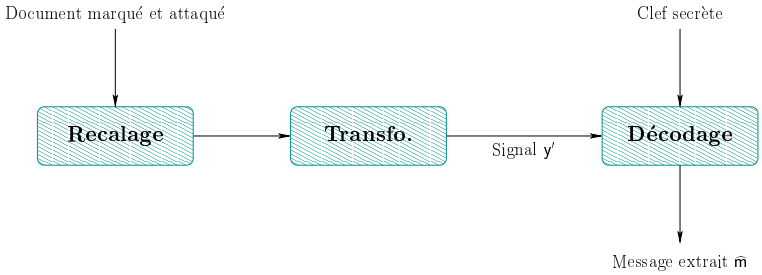


FIG. 3.1 – Schéma d'extraction classique : le document est recalé géométriquement avant extraction

modèles utilisés sont limités : il est impossible de prévoir toutes les transformations géométriques susceptibles d'être appliquées (c'est un compromis entre complexité du modèle et diversité des attaques prises en compte). Or, les attaques les plus malicieuses, comme celle de Stirmark (figure 3.2), utilisent des transformations locales difficiles à modéliser. Il en résulte des défauts de recalage introduisant des pertes de performance.

Quelques études ont tenté de modéliser l'impact des défauts de synchronisation sur la performance de schémas de tatouage. Une des premières fut celle d'Eggers *et al.* [EBG02], auteurs d'un schéma de tatouage basé sur la quantification [EG00, EBTG02]. Ce type de schéma est sensible aux facteurs d'échelle : les pas de quantification sont alors décalés. Ils proposent d'introduire des séquences pilotes et étudient l'impact d'erreurs de synchronisation sur leur schéma. Baudry *et al.* [BNM02] proposent une technique originale basée sur une modélisation markovienne des désynchronisations. Enfin, un article récent [LOJPG03] étudie l'influence des désynchronisations (*random jitter attack*) sur les performances (taux d'erreurs par bit) d'un schéma de tatouage par étalement de spectre dans le domaine spatial. Les désynchronisations sont modélisées par un ajout de bruit gaussien dépendant du signal marqué.

Nous allons étudier dans cette section l'impact de ces erreurs en introduisant dans notre modèle de tatouage un facteur de désynchronisation. Nous verrons les améliorations apportées en appliquant les résultats sur le marquage de signaux issus d'une transformée en ondelettes.

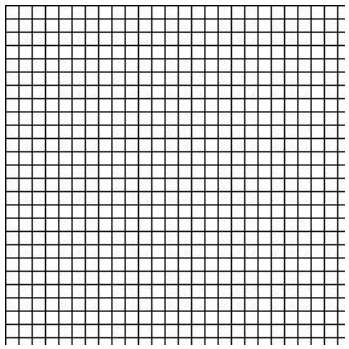
3.1.1 Modélisation

Nous notons le signal discret y sous une forme continue $y(t) = \sum_{i=1}^{m} y_i \times \mathrm{sinc}^d(t-i)$, avec d dimension du signal ($d = 1$ pour un signal audio, 2 pour des images, ...). Nous pouvons en déduire que si l'on introduit une erreur de synchronisation Δ, le signal discret y' issu de l'échantillonnage de $y(t+\Delta)$ est donné par

$$y'_i = c_i \times y_i + n_i, \tag{3.1}$$

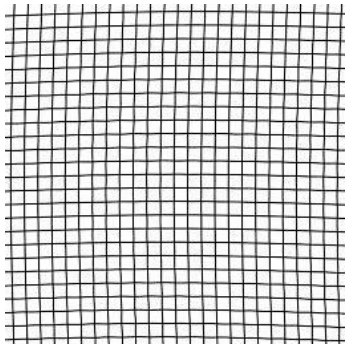
avec $c_i = \mathrm{sinc}^d(\Delta)$ et n_i bruit de variance $\sigma^2_{N_i} = \left(1 - c_i^2\right) \sigma^2_{Y_i}$ représentant le bruit de l'interférence des échantillons voisins. Plus généralement, nous exprimons cette variance

(a) Image marquée y : *Lena*



(b) Image marquée x : grille de 512×512 pixels

(c) Image attaquée y'



(d) Image attaquée y'

FIG. 3.2 – Effet de la distorsion géométrique locale appliquée par Stirmark. Exemples tirés du site de F. Petitcolas (`http ://www.petitcolas.net/`)

par $\sigma^2_{N_i} = a_i \times \sigma^2_{Y_i}$. Une erreur de $\Delta = 1$ suffit à annuler complètement la réponse de la marque. De ce fait, on considère que $c_i = 0$ pour $\Delta \geq 1$. Le rapport signal-à-bruit atteignable en prenant en compte l'information adjacente et cette désynchronisation est exprimé par

$$\frac{E_b}{N_0} = \frac{c_i^2 \times \gamma_i^2 \sigma^2_{W_i}}{a_i \times \gamma_i^2 \left(\sigma^2_{X_i} + \sigma^2_{W_i}\right) + \sigma^2_{Z_i}}, \tag{3.2}$$

obtenu en reprenant la maximisation de la section 1.2. Cette nouvelle mesure de performance nous permet de reprendre le jeu entre attaquant et défenseur afin de trouver la stratégie d'insertion optimale. On remarque que pour le cas $c_i = a_i = 1$, on retrouve la mesure de performance utilisée dans la partie 2 (sans prise en compte de l'information adjacente), et que pour $a_i = 0$, on tombe sur celle la section 1 de cette partie. Cette mesure peut être vue comme une généralisation des cas étudiés précédemment.

3.1.2 Résolution du jeu

Les mesures de distorsions restent inchangées par rapport au jeu défini dans le premier chapitre :

$$D_{\mathsf{xy}} = \frac{1}{m} \sum_{i=1}^{m} \varphi_i^2 \left[\sigma^2_{X_i} (1 - \bar{\gamma}_i)^2 + \bar{\gamma}_i^2 \sigma^2_{W_i}\right] \tag{3.3}$$

$$D_{\mathsf{xy}'} = \frac{1}{m} \sum_{i=1}^{m} \varphi_i^2 \left[\sigma^2_{X_i} (1 - \bar{\gamma}_i \bar{\bar{\gamma}}_i)^2 + \bar{\gamma}_i^2 \bar{\bar{\gamma}}_i^2 \sigma^2_{W_i} + \sigma^2_{Z_i}\right]. \tag{3.4}$$

Elles ne prennent pas en compte la distorsion visuelle introduite par le décalage. Ce type de mesure reste un problème ouvert et ne pourrait être pris en compte par une simple pondération φ_i.

Attaque

La résolution de la minimisation lagrangienne est très similaire aux cas vus précédemment. La mise à zéro des dérivées de la fonctionnelle J^i_λ par rapport à γ_i et $\sigma^2_{Z_i}$ donne les paramètres

$$\gamma_i^{\mathrm{a}} = \frac{1}{1 - a_i} \left[\gamma_i^{\mathrm{w}} - \frac{c_i \times \sigma_{W_i}}{\sqrt{\lambda} \varphi_i \left(\sigma^2_{X_i} + \sigma^2_{W_i}\right)}\right] \tag{3.5}$$

$$\sigma^{\mathrm{a}}_{Z_i} = \sqrt{\gamma_i^{\mathrm{a}} \left(\gamma_i^{\mathrm{w}} - \gamma_i^{\mathrm{a}}\right) \left(\sigma^2_{X_i} + \sigma^2_{W_i}\right)}, \tag{3.6}$$

optimaux pour $\sigma^{\mathrm{a}}_{Z_i} \geq 0$ (domaine $\mathcal{D}_2$). L'exploration du cas limite $\sigma_{Z_i} = 0$ donne deux autres stratégies d'attaque : annulation du signal ($\gamma_i = 0$) si $\sigma_{W_i} < \sqrt{\lambda} \varphi_i \sigma^2_{X_i} / c_i$ (domaine $\mathcal{D}_1$), et filtrage de Wiener ($\gamma_i = \gamma_i^{\mathrm{w}}$) si

$$\gamma_i^{\mathrm{a}} > \gamma_i^{\mathrm{w}} \quad \Rightarrow \quad \frac{1}{1 - a_i} \left[\sigma_{W_i} - \sqrt{\lambda} \varphi_i \sigma^2_{X_i}\right] + \sigma^2_{W_i} < 0, \tag{3.7}$$

définissant le domaine $\mathcal{D}_3$. Les domaines définissant les stratégies et donc l'attaque optimale dépendent donc des paramètres de désynchronisation.

Défense

En considérant la stratégie d'attaque optimale, nous recherchons une méthode de défense maximisant la performance E_b/N_0. Nous calculons pour les trois stratégies une réponse adaptée, en veillant à rester dans les limites des domaines pour lesquels elles sont définies. La stratégie adaptée au domaine $\mathcal{D}_2$ (paramètres d'attaque des équations 3.5 et 3.6) s'avère être la plus performante de toutes. Elle est donnée par

$$\begin{aligned}\sigma^{\star}_{W_i} &= \frac{\varphi_i^2(\lambda-\chi(1-a_i))\sigma^2_{X_i} - c_i^2 + \sqrt{\left(\varphi_i^2(\lambda-\chi(1-a_i))\sigma^2_{X_i} - c_i^2\right)^2 + 4\lambda\varphi_i^2\sigma^2_{X_i}c_i^2}}{2c_i \times \sqrt{\lambda}\varphi_i} \\ &\quad \text{si } \lambda > \chi \text{ ou } \sigma_{X_i} < \frac{c_i}{\varphi_i\sqrt{a_i\,(\chi-\lambda)}} \\ &= 0 \text{ sinon.}\end{aligned} \tag{3.8}$$

La défense est très similaire à celle trouvée pour le chapitre 1 (on retrouve la même formule si l'on pose $a_i = 0$ et $c_i = 1$). Elle prend en compte les paramètres de désynchronisation. On observe néanmoins une limite au delà de laquelle les échantillons ne sont plus marqués, comme pour l'optimisation concernant le tatouage par étalement de spectre classique (chapitre 3 de la seconde partie). On voit également que le seuil décroît avec c_i, c'est-à-dire que le nombre d'échantillons marqués sera d'autant plus faible que l'erreur de recalage Δ est forte : il est inutile de marquer les coefficients qui seront mal synchronisés et participerons peu à la performance de la transmission. Enfin, on montre que l'extracteur optimal est $\beta_i^{\star} \propto \varphi_i$ si on utilise les stratégies que nous venons de montrer.

3.1.3 Résultats : application au tatouage d'image utilisant la transformée en ondelettes

La transformée en ondelettes est une décomposition multi-résolution : elle est appliquée récursivement sur une version de plus en plus réduite du signal. De ce fait, une erreur de synchronisation de Δ au niveau spatial entraîne un décalage de $\Delta/2$ dans les sous-bandes issues du premier niveau de décomposition, et une erreur de $\Delta/4$ pour celles du second (voir la figure 3.3). Les sous-bandes de plus basses fréquences offrent une résistance accrue vis-à-vis de la désynchronisation. Nous exploitons cette propriété dans les résultats présentés ici, appliqués aux images en niveaux de gris. Pour un décalage de Δ pixels dans le domaine spatial, les paramètres c_i (pondération mesurant énergie exploitable) et a_i (pondération sur l'interférence des échantillons voisins) pour les sous-bandes issues d'un niveau de décomposition d sont donnés par

$$\Delta_i = \frac{\Delta}{2^d} \tag{3.9}$$

$$\begin{aligned} c_i &= \operatorname{sinc}^2(\Delta_i) \text{ si } \Delta_i < 1 \\ &= 0 \text{ sinon} \qquad (3.10) \\ a_i &= 1 - c_i^2. \qquad (3.11) \end{aligned}$$

La figure 3.4 confirme la justesse de ce choix : la réponse réelle d'une marque d'énergie 1 calculée en fonction du décalage (courbes en traits pleins) est proche[1] de la fonction $\operatorname{sinc}^2()$ (traits pointillés) pour des valeurs de $\Delta_i < 1$. À chaque échantillon x_i est donc associé un couple (a_i, c_i) dépendant de la résolution dont est issu l'échantillon considéré. Le mode opératoire utilisé dans ces expérimentations est le même que ceux vus dans les résultats précédents : le signal hôte x est obtenu par transformée en ondelettes de l'image à marquer, la marque est ajoutée en utilisant les paramètres de défense optimaux vus ci-dessus puis l'attaque optimale est appliquée. La distorsion $D_{\mathsf{xy'}}$ obtenue est notée en abscisse et le rapport signal-à-bruit (équation 3.2) en ordonnée.

La figure 3.5 montre l'impact du décalage sur la performance du tatouage lorsque l'on utilise une DWT sur trois niveaux. Il est alors possible de résister (rapport signal-à-bruit supérieur à zéro) à un décalage maximal de $2^3 = 8$ pixels. La désynchronisation introduit une forte perte de performance : pour l'image *Lena*, on passe d'une capacité maximale totale de 3850 bits[2] (pour $\Delta = 0$) à 24 bits ($\Delta = 4$ pixels) pour un PSNR de 32 dB entre image originale et image marquée (soit une distorsion $D_{\mathsf{xy'}} = 40$). En utilisant une DWT sur cinq niveaux (figure 3.6), la perte est plus faible (capacité totale d'environ 150 bits avec le même niveau d'attaque et un décalage identique). Il est donc préférable d'utiliser une DWT avec un nombre important de décompositions pour prévenir les désynchronisations. Cela est confirmé par la figure 3.7. La formule de $\sigma^\star_{W_i}$ (équation 3.8) fait que les sous-bandes de plus faibles résolutions seront privilégiées à l'insertion. La figure 3.8 indique la participation de chacune des sous-bandes dans le rapport signal-à-bruit global : la participation des sous-bandes de plus basses fréquences devient prédominante pour les attaque importantes, malgré leurs faibles nombres d'échantillons.

3.2 Utilisation de la réalisation du signal

Les optimisations par théorie des jeux faites dans les chapitres précédents s'appuient sur une connaissance commune entre les deux parties : la distribution statistique du signal hôte (c'est-à-dire l'ensemble $\{\sigma^2_{X_1}, \sigma^2_{X_2}, \ldots, \sigma^2_{X_m}\}$) et le facteur de pondération perceptuelle φ_i. Les paramètres des stratégies d'attaque et de défense sont exprimés en fonction de cette connaissance. Or l'attaquant dispose en plus de la réalisation y du signal marqué. Il peut donc estimer plus finement la distribution de X :

$$\Pr\left(X_i = x_i \,|\, Y_i = y_i\right) = \frac{\Pr(Y_i = y_i \,|\, X_i = x_i) \times \Pr(X_i = x_i)}{\Pr(Y_i = y_i)}$$

[1] La légère différence vient probablement du fait du support fini des filtres utilisés pour la transformée.

[2] Estimée par la formule $\mathcal{C} = \frac{m}{2} \log_2 \left[1 + \frac{E_b}{m \times N_0}\right]$.

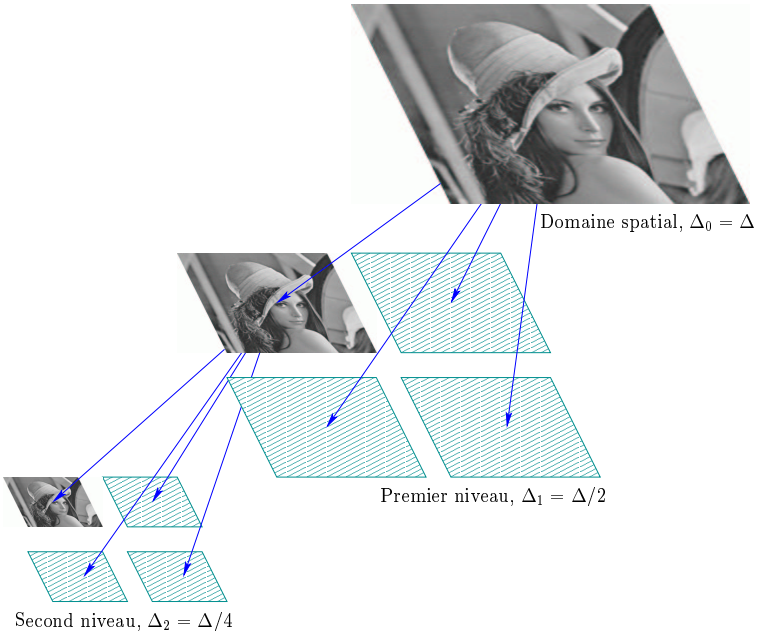


FIG. 3.3 – Pyramide formée par une transformée en ondelettes. L'erreur de synchronisation se réduit au fur et à mesure de la décomposition

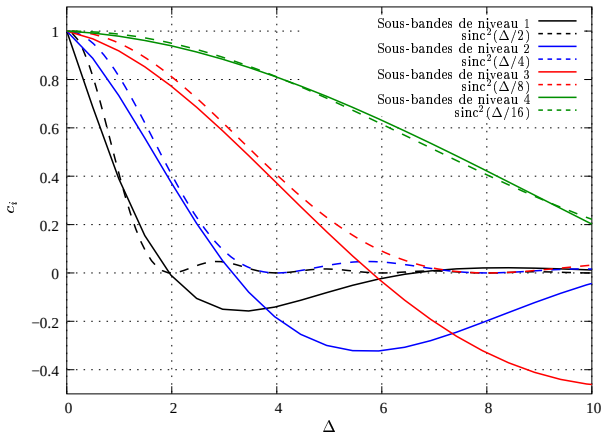
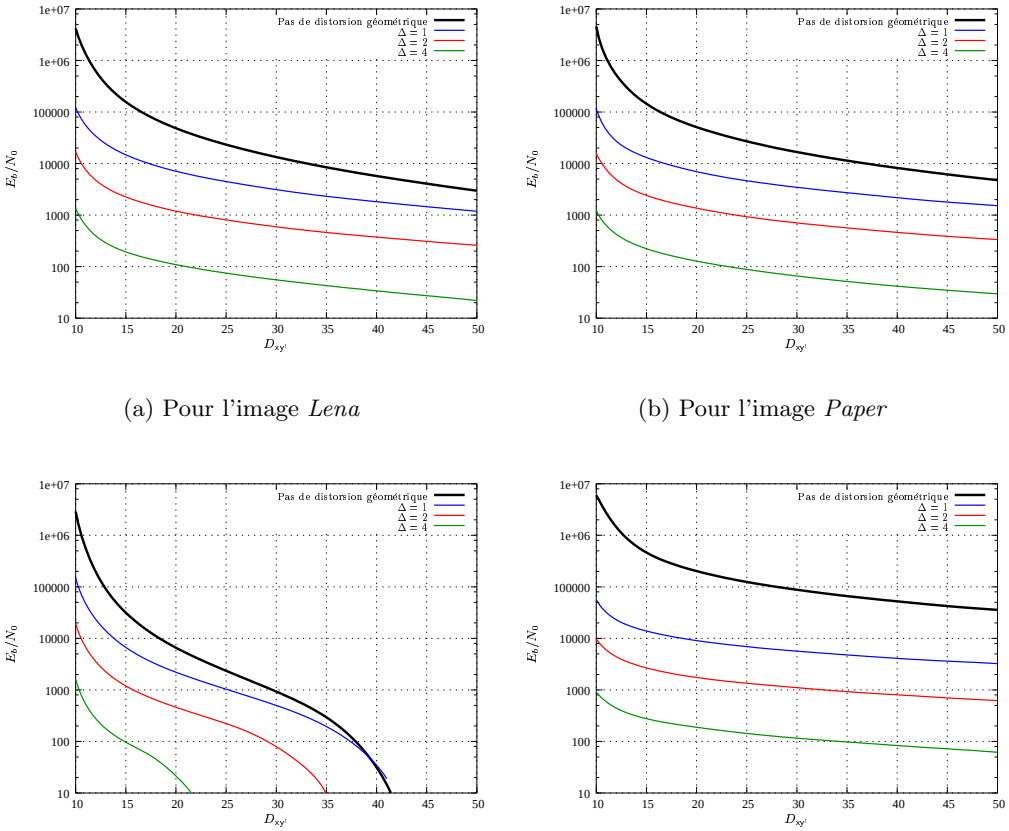


FIG. 3.4 – Réponses d'une marque d'énergie 1 dans plusieurs niveaux de résolution en fonction de l'erreur de recalage Δ



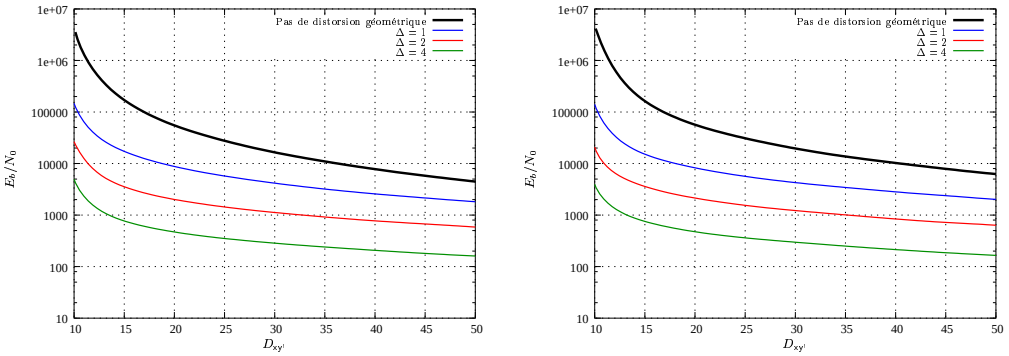
(a) Pour l'image *Lena*

(b) Pour l'image *Paper*

(c) Pour l'image *Rose*

(d) Pour l'image *Baboon*

FIG. 3.5 – Impact d'un recalage d'une imprécision de Δ pixels, avec une DWT sur 3 niveaux. Les performances données sont celles obtenues face à l'attaque optimale, avec une insertion telle que $D_{\mathsf{xy}} = 10$ ($\varphi_i = 1$)



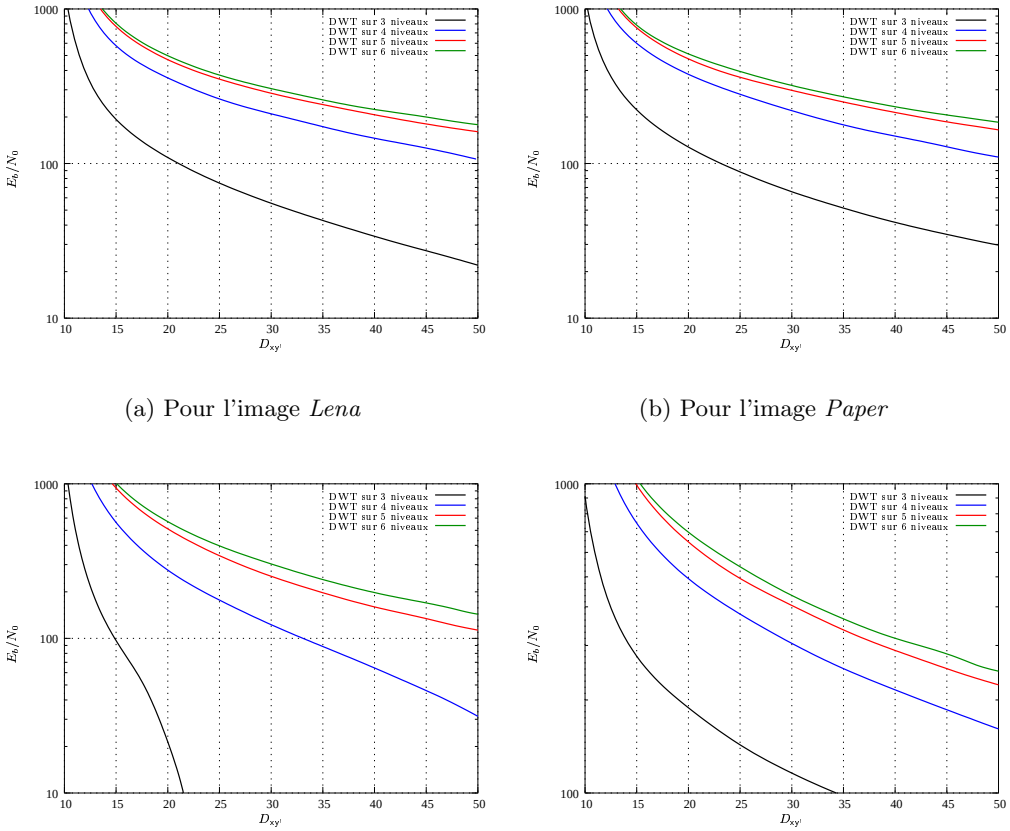
(a) Pour l'image *Lena*

(b) Pour l'image *Paper*

(c) Pour l'image *Rose*

(d) Pour l'image *Baboon*

FIG. 3.6 – Impact d'un recalage d'une imprécision de Δ pixels, avec une DWT sur 5 niveaux. Les performances données sont celles obtenues face à l'attaque optimale, avec une insertion telle que $D_{\mathsf{xy}} = 10$ ($\varphi_i = 1$)



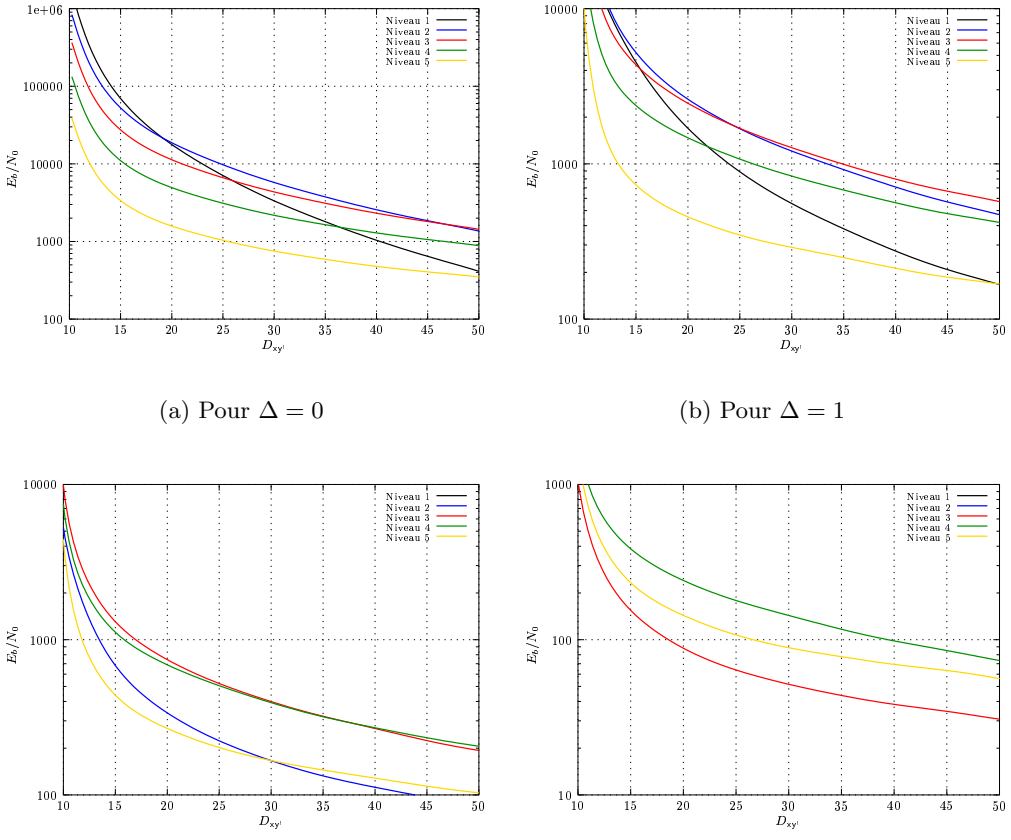
(a) Pour l'image *Lena*

(b) Pour l'image *Paper*

(c) Pour l'image *Rose*

(d) Pour l'image *Baboon*

FIG. 3.7 – Performances obtenues selon le niveau de décomposition en ondelettes choisi, avec une erreur de recalage de $\Delta = 4$ pixels. Les performances données sont celles obtenues face à l'attaque optimale, avec une insertion telle que $D_{\mathsf{xy}} = 10$ ($\varphi_i = 1$)



(a) Pour $\Delta = 0$

(b) Pour $\Delta = 1$

(c) Pour $\Delta = 2$

(d) Pour $\Delta = 4$

FIG. 3.8 – Contribution de chaque niveau de résolution dans la performance du schéma, en fonction de l'erreur de synchronisation. Les performances données sont celles obtenues face à l'attaque optimale, avec une insertion telle que $D_{\mathsf{xy}} = 10$ ($\varphi_i = 1$)

$$\sim \quad \mathcal{N}\left(\frac{\gamma_i^{\mathrm{W}} y_i}{\bar{\gamma}_i}, \gamma_i^{\mathrm{W}} \sigma_{W_i}^2\right), \tag{3.12}$$

avec γ_i^{W} coefficient multiplicateur d'un filtre de Wiener, comme vu précédemment. En prenant cette estimation de x, la formule de distorsion $D_{\mathsf{xy}'}$ devient

$$\begin{aligned} D_{\mathsf{xy}'} &= \frac{1}{m}\mathbb{E}\left[\sum_{i=1}^{m} \varphi_i^2 \left(X_i - Y_i'\right)^2\right] \\ &= \frac{1}{m}\sum_{i=1}^{m} \varphi_i^2 \left[\mathbb{E}\left[X_i^2\right] + \bar{\bar{\gamma}}_i^2 \mathbb{E}\left[Y_i^2\right] - 2\bar{\bar{\gamma}}_i \mathbb{E}\left[X_i Y_i\right] + \mathbb{E}\left[Z_i\right]\right] \\ &= \frac{1}{m}\sum_{i=1}^{m} \varphi_i^2 \left[y_i^2 \left(\frac{\gamma_i^{\mathrm{W}}}{\bar{\gamma}_i} - \bar{\bar{\gamma}}_i\right)^2 + \gamma_i^{\mathrm{W}} \sigma_{W_i}^2 + \sigma_{Z_i}^2\right]. \end{aligned} \tag{3.13}$$

Nous reprenons l'optimisation par la théorie du jeu afin de définir une stratégie d'attaque et la défense correspondante. Le tatouage avec prise en compte de l'information adjacente utilise la réalisation de X pour définir le signal transmis (chapitre 1). Ce type de technique est souvent appelé tatouage informé [MCB00, EG02]. Par analogie, nous qualifions l'attaque prenant en compte la réalisation de Y d'**attaque informée**.

3.2.1 Attaque informée

Nous utilisons une formulation lagrangienne pour rechercher la stratégie minimisant le rapport E_b/N_0. De par la forme additive de la fonctionnelle, l'optimisation peut se faire échantillon par échantillon :

$$\left(\bar{\bar{\gamma}}_i^{\star}, \sigma_{Z_i}^{\star}\right) = \arg \min_{\bar{\bar{\gamma}}_i, \sigma_{Z_i} \geq 0} \left\{ J_\lambda^i = \frac{\gamma_i^2 \sigma_{W_i}^2}{\sigma_{Z_i}^2} + \lambda \varphi_i^2 \left[y_i \left(\frac{\gamma_i^{\mathrm{W}}}{\bar{\gamma}_i} - \bar{\bar{\gamma}}_i\right)^2 + \gamma_i^{\mathrm{W}} \sigma_{W_i}^2 + \sigma_{Z_i}^2 \right] \right\}. \tag{3.14}$$

Nous résolvons d'abord pour le cas général par annulation des dérivées, puis nous rechercherons une solution aux bords du domaine ($\sigma_{Z_i} = 0$). L'annulation des dérivées donne

$$\frac{\partial J_\lambda^i}{\partial \bar{\bar{\gamma}}_i} = 0 \quad \Leftrightarrow \quad \bar{\bar{\gamma}}\left[\frac{\bar{\gamma}_i^2 \sigma_{W_i}^2}{\sigma_{Z_i}^2} + \lambda \varphi_i y_i^2\right] = \lambda \varphi_i^2 y_i^2 \frac{\gamma_i^{\mathrm{W}}}{\bar{\gamma}_i} \tag{3.15}$$

$$\frac{\partial J_\lambda^i}{\partial \sigma_{Z_i}^2} = 0 \quad \Leftrightarrow \quad \lambda \varphi_i^2 = \frac{\gamma_i^2 \sigma_{W_i}^2}{\sigma_{Z_i}^4}, \tag{3.16}$$

et fournit les paramètres solution

$$\bar{\bar{\gamma}}_i^{\mathrm{a}} = \frac{\gamma_i^{\mathrm{W}}}{\bar{\gamma}_i} - \frac{\bar{\gamma}_i \sigma_{W_i}}{\sqrt{\lambda} \varphi_i \times y_i^2} \tag{3.17}$$

$$\sigma_{Z_i}^{\mathrm{a}} = \sqrt{y_i^2 \bar{\bar{\gamma}}_i \left(\frac{\gamma_i^{\mathrm{W}}}{\bar{\gamma}_i} - \bar{\bar{\gamma}}_i\right)}. \tag{3.18}$$

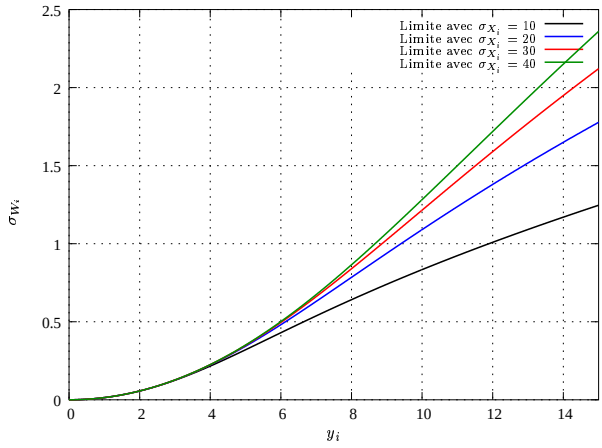


FIG. 3.9 – Limites entre les deux domaines d'attaque, définies par l'équation 3.9

Ce couple définit la stratégie $\mathsf{a_I}(i) = (\bar{\gamma}_i^{\mathrm{a}}, \sigma_{Z_i}^{\mathrm{a}})$. En posant $y_i^2/\bar{\gamma}_i^2 \simeq \sigma_{X_i}^2 + \sigma_{W_i}^2$, on retrouve les formules des stratégies déterminés dans les chapitres 3 (partie 2) et 1 (partie 3). La solution n'est valable que pour

$$\sigma_{W_i} \leq \frac{\sqrt{\lambda}\varphi_i\gamma_i^{\mathrm{w}}}{\bar{\gamma}_i^2} y_i^2. \tag{3.19}$$

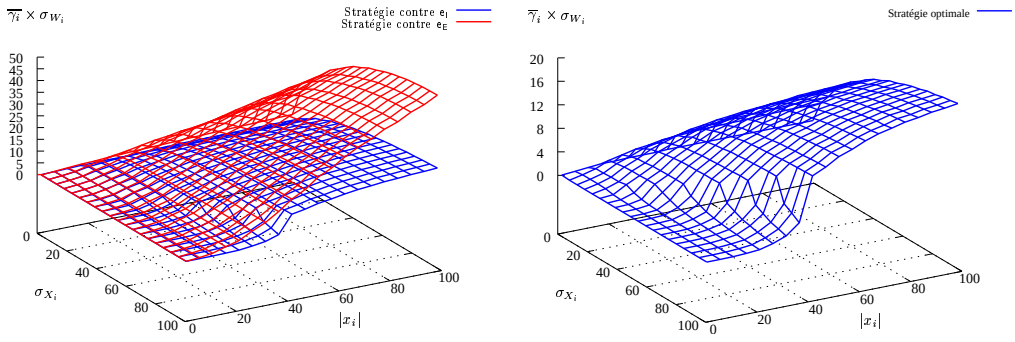
Le domaine $\mathcal{D}_2$ est défini par le respect de cette contrainte. Le reste du domaine de validité est noté $\mathcal{D}_1$. La figure 3.9 montre quelques exemples de limite entre les deux domaines. L'exploration du cas $\sigma_{Z_i} = 0$ nous donne la seconde stratégie $\mathsf{a_E}(i) = (0, 0)$. Grâce à des calculs similaires à ceux présentés par l'annexe A.2.2, on montre que la stratégie $\mathsf{a_E}$ est optimale sur $\mathcal{D}_1$ et $\mathsf{a_I}$ sur $\mathcal{D}_2$.

La forme de l'attaque informée est très proche de celles vues dans les chapitres précédents. Les échantillons dont la réalisation est faible par rapport à l'énergie de la marque ajoutée sont annulés : la distorsion introduite est alors peu importante. Les autres sont bruités puis filtrés.

3.2.2 Stratégie de défense

La stratégie de défense optimale est obtenue par maximisation de $J_\lambda + \chi D_{\mathsf{xy}}$. Or la fonctionnelle J_λ est exprimée en fonction de y et donc de w. Ce dernier est inconnu avant la projection car la prise en compte de l'information adjacente fait qu'il dépend de la stratégie de défense. Il est donc impossible de résoudre directement l'optimisation. L'estimation de y_i^2 par $\bar{\gamma}_i^2\left(\sigma_{X_i}^2 + \sigma_{W_i}^2\right)$ nous ramène à l'optimisation déjà résolue au chapitre 1 et donc à la stratégie de l'équation 1.26 (page 110).

Un autre point de vue est, à l'image de l'attaque, de prendre en compte la réalisation de X dans la mesure de distorsion, et d'estimer y_i^2 par $\bar{\gamma}_i^2\left(x_i^2 + \sigma_{W_i}^2\right)$. La distorsion



(a) Stratégies d'insertion contre l'attaque par effacement et l'attaque intermédiaire

(b) Stratégie globale, obtenue en sélectionnant la meilleure valeur de J_χ^i

FIG. 3.10 – Stratégie d'insertion optimale obtenue en posant $y_i^2 \simeq \bar{\gamma}_i^2 \left(x_i^2 + \sigma_{W_i}^2\right)$ ($\lambda = 0,0001$, $\chi = 0,0005$ et $\varphi_i = 1$)

d'insertion devient

$$D_{\text{xy}} = \frac{1}{m} \sum_{i=1}^{m} \varphi_i^2 \left[x_i^2 \left(1 - \bar{\gamma}_i\right)^2 + \bar{\gamma}_i^2 \sigma_{W_i}^2 \right]. \tag{3.20}$$

La minimisation de la fonctionnelle $J_\chi = J_\lambda + \chi D_{\text{xy}}$ donne, quelle que soit la stratégie parmi les deux exposées à la section ci-dessus, la forme de filtrage optimale suivante :

$$\bar{\gamma}_i^\star = \frac{x_i^2}{x_i^2 + \sigma_{W_i}^2}. \tag{3.21}$$

Rechercher la valeur de σ_{W_i} optimale est plus problématique. Dans le domaine $\mathcal{D}_1$ (la stratégie d'attaque optimale est l'annulation), la dérivée de J_χ^i est négative et le maximum est obtenu au bord du domaine (équation 3.19). La solution est donc la racine du polynôme du troisième degré

$$\sigma_{W_i}^3 + \sigma_{W_i} \sigma_{X_i}^2 - \sigma_{W_i}^2 \left(\sqrt{\lambda} \varphi_i \sigma_{X_i}^2\right) - x_i^2 \sqrt{\lambda} \varphi_i \sigma_{X_i}^2. \tag{3.22}$$

La maximisation pour le domaine $\mathcal{D}_2$ est probablement impossible à résoudre analytiquement. En utilisant une recherche numérique du maximum, nous obtenons le résultat présenté par la courbe bleue de la figure 3.10(a). La stratégie globale optimale est obtenue en choisissant la stratégie donnant la meilleure valeur de J_χ^i (figure 3.10(b)).

3.2.3 Résultats

Les figures 3.11 à 3.13 montre l'impact de l'attaque informée sur les performances de transmission. Il est comparé aux attaques précédemment testées : ajout de bruit gaussien uniforme, de bruit proportionnel à l'énergie de la marque et attaque optimale

vue dans le chapitre 1 de cette partie (attaque non informée). Trois stratégies ont été utilisées : énergie de la marque uniforme, énergie de type PSC et stratégie de défense optimale (chapitre 1).

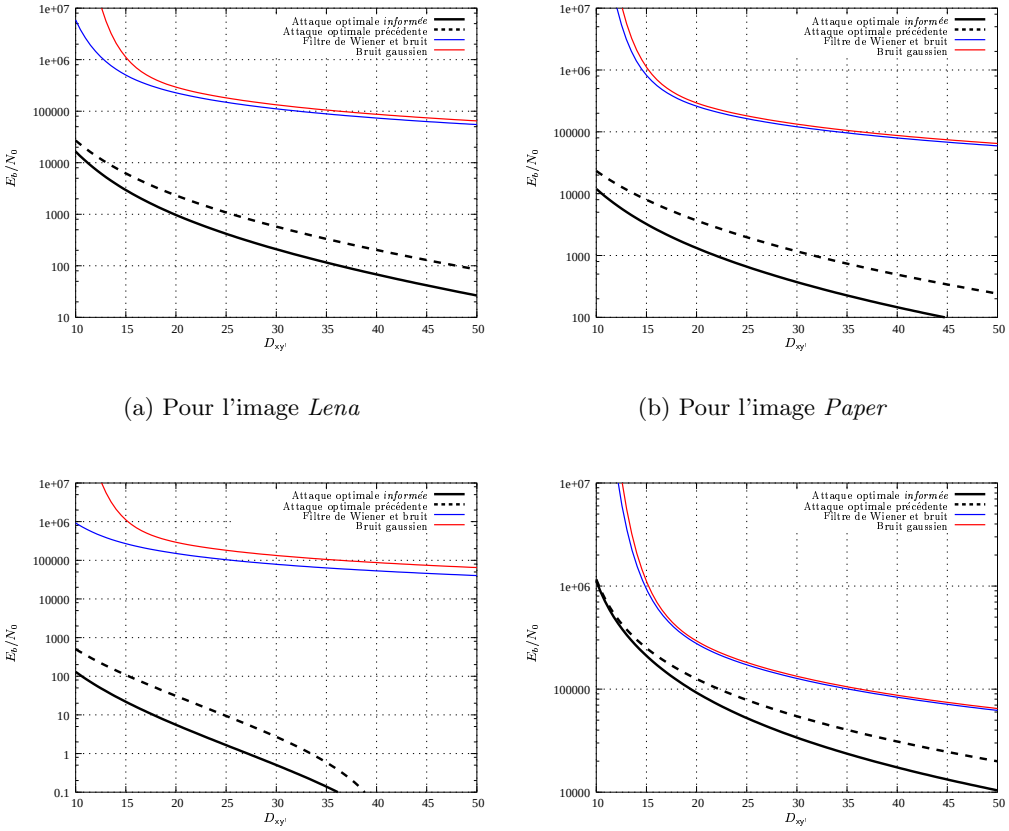
La prise en compte de la réalisation de Y lors de l'attaque apporte un gain (perte de performance plus importante) quelle que soit la technique d'insertion choisie. En utilisant une mesure de distorsion plus fine, l'attaque est plus efficace. Alors que la stratégie de défense du premier chapitre nous assurait une capacité totale de 6200 bits sur l'image *Paper* face à l'attaque non informée de distorsion $D_{\mathsf{xy}'} = 40$ (PSNR d'environ 32 dB), l'attaque informée fait chuter ce chiffre à 2310 bits. La différence est plus limitée pour les distorsions d'attaque plus faibles.

La figure 3.14 montre les performances atteignables face à l'attaque optimale informée avec plusieurs répartitions de l'énergie de la marque. Sans surprise, les stratégies définies par max-min sont les plus performantes, et la stratégie que nous venons de voir, avec prise en compte de la réalisation de X, est meilleure que celle du chapitre 1. En utilisant ses nouvelles stratégies, les performances atteintes restent inférieures à celles vues dans le chapitre 1 (courbes en pointillés). Néanmoins, la stratégie de défense a été développée sur la base d'une approximation de y_i^2. On peut supposer qu'une estimation plus précise améliorerait les résultats pour se rapprocher des performances précédentes.

Conclusion

La première partie de ce chapitre nous a permis de justifier une idée déjà employée empiriquement [ZL02] : privilégier les plus basses fréquences du document à marquer permet d'obtenir une meilleure résistance vis-à-vis des attaques géométriques. Du fait de sa construction sous la forme d'une pyramide de sous-bandes de différentes résolutions, la transformée en ondelettes est un outil de choix pour appliquer nos résultats. Plus le décalage prévu est important et plus l'énergie de la marque sera localisée dans les sous-bandes de plus faibles résolutions.

La seconde partie a vu la définition d'une forme d'attaque plus performante que celle développée dans le chapitre 1. Elle utilise le fait que l'attaquant connaît parfaitement le signal qu'il reçoit et de ce fait a été baptisée attaque informée. L'attaquant peut donc estimer plus finement la distorsion qu'il introduit et être plus efficace. Les résultats montrent un meilleur comportement que l'attaque classique vue précédemment. La stratégie de défense correspondante n'a pu être calculée analytiquement et nous avons utilisé pour nos résultats une résolution numérique. De plus, nous sommes obligés de passer par une estimation de y afin de prévoir la réaction de l'attaque dans notre optimisation. Malgré cela, cette défense apporte une amélioration face à l'attaque informée.



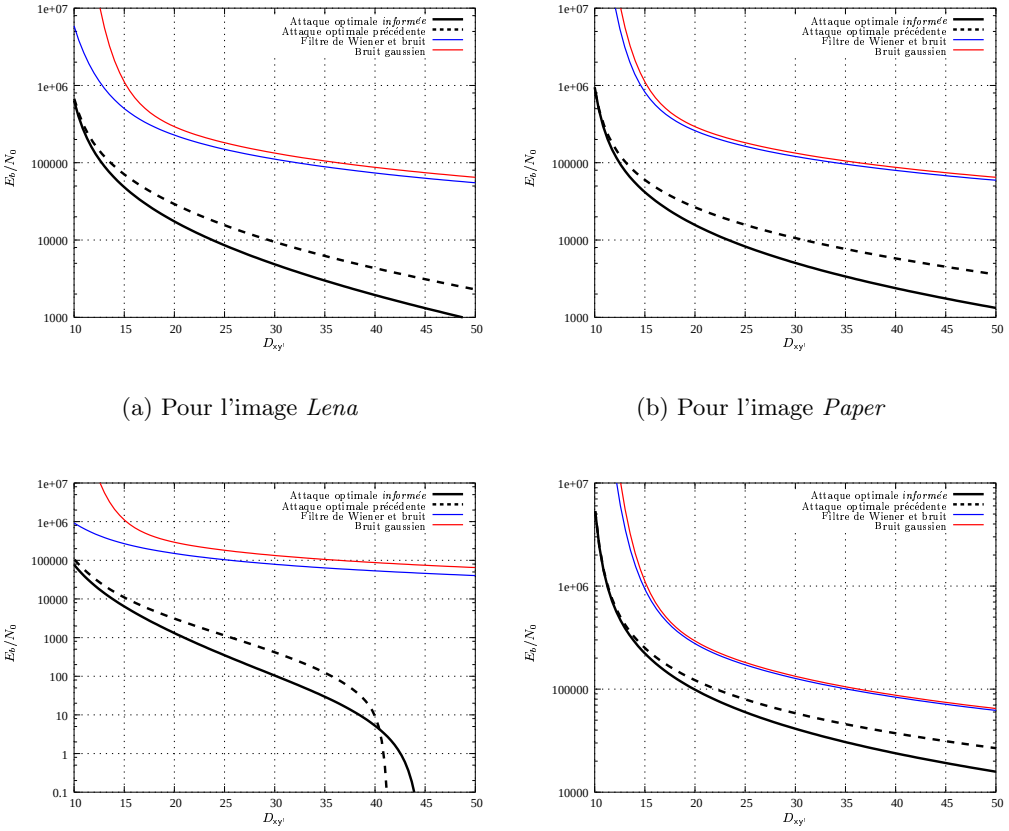
(a) Pour l'image *Lena*

(b) Pour l'image *Paper*

(c) Pour l'image *Rose*

(d) Pour l'image *Baboon*

FIG. 3.11 – Apport de l'attaque avec prise en compte de la réalisation de Y (attaque informée) vis-à-vis des attaques vues précédemment. L'énergie de la marque est constante (répartition uniforme) telle que $D_{\mathsf{xy}} = 10$ ($\varphi_i = 1$)



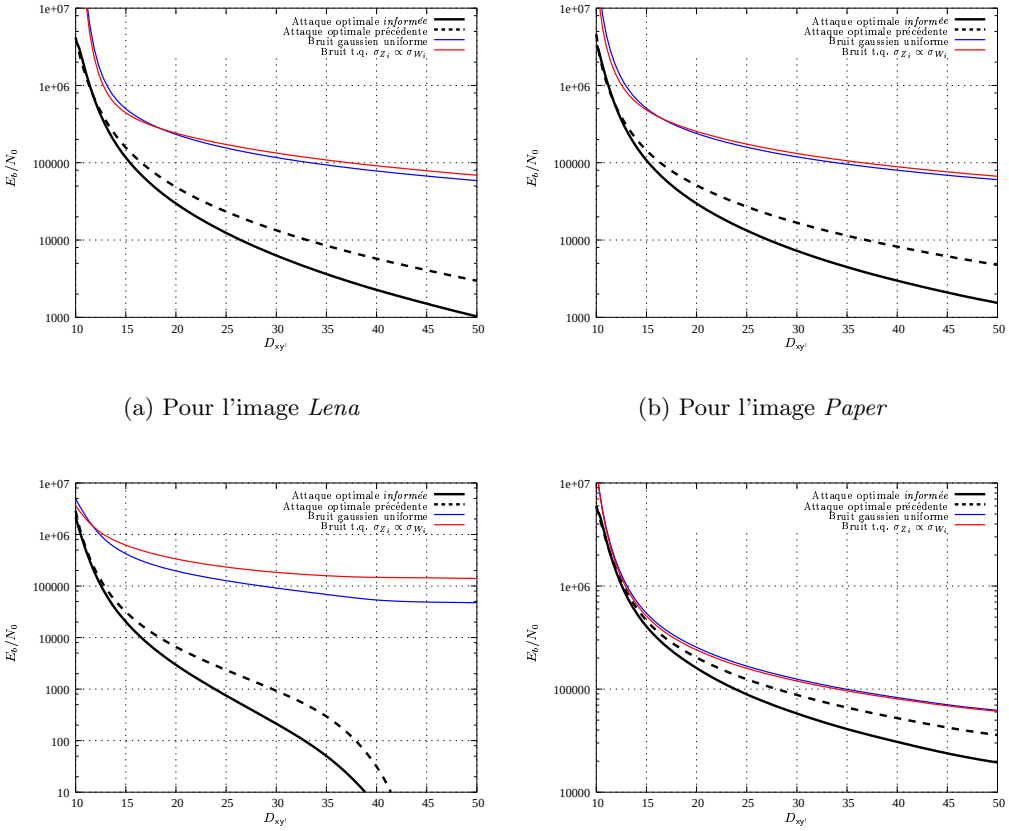
(a) Pour l'image *Lena*

(b) Pour l'image *Paper*

(c) Pour l'image *Rose*

(d) Pour l'image *Baboon*

FIG. 3.12 – Apport de l'attaque avec prise en compte de la réalisation de Y (attaque informée) vis-à-vis des attaques vues précédemment. L'énergie de la marque est du type $\sigma_{W_i} \propto \sigma_{X_i}$ telle que $D_{\mathsf{xy}} = 10$ ($\varphi_i = 1$)



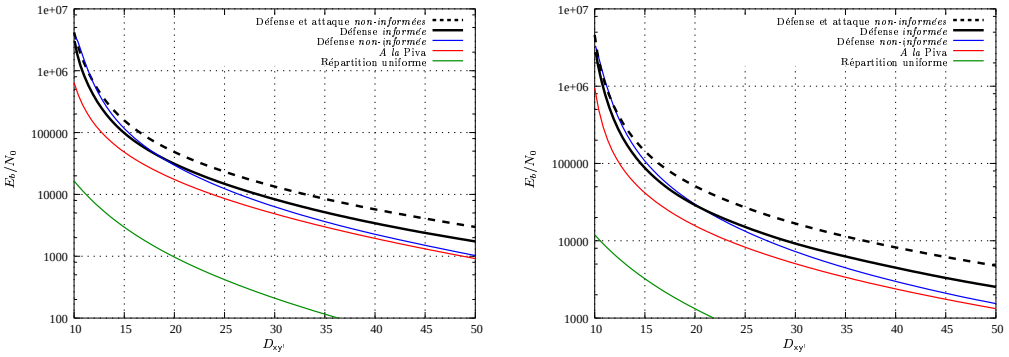
(a) Pour l'image *Lena*

(b) Pour l'image *Paper*

(c) Pour l'image *Rose*

(d) Pour l'image *Baboon*

FIG. 3.13 – Apport de l'attaque avec prise en compte de la réalisation de Y (attaque informée) vis-à-vis des attaques vues précédemment. L'énergie est définie par la stratégie de défense vue dans le chapitre 1 de cette partie, et telle que $D_{\mathsf{xy}} = 10$ ($\varphi_i = 1$)



(a) Pour l'image *Lena*

(b) Pour l'image *Paper*

(c) Pour l'image *Rose*

(d) Pour l'image *Baboon*

FIG. 3.14 – Performance de plusieurs répartitions de l'énergie de la marque face à l'attaque optimale prenant en compte la réalisation de Y (attaque informée). Les performances des stratégies vues au chapitre 1 (en pointillés) sont mises à titre de comparaison ($D_{\mathsf{xy}} = 10$ et $\varphi_i = 1$)

Chapitre 4

Mise en pratique

Notre schéma de tatouage est décomposable en deux parties : l'étalement de spectre, permettant de répartir l'énergie de la marque en fonction d'une distorsion d'insertion et d'attaque (avec des mesures de distorsion utilisant éventuellement une pondération perceptuelle), et le codage du message qui utilise un dictionnaire basé sur les travaux de Costa [Cos83]. Nous avons montré dans les chapitres précédents les résultats de chacune de ses parties, c'est-à-dire le rapport signal-à-bruit que l'on peut atteindre en utilisant nos stratégies d'insertion et les probabilités d'erreur de notre technique de codage. Or, le lien entre ces deux aspects n'a pas encore été fait.

Nous proposons dans ce dernier chapitre de voir comment appliquer nos travaux afin de proposer une chaîne de tatouage complète. Nous utilisons ici des images monochromes avec une transformée en ondelettes. Ce domaine est particulièrement bien adapté, à la fois pour ses propriétés statistiques (bon pouvoir de décorrélation) et pour sa bonne adéquation avec la prise en compte de désynchronisations géométriques, comme vu dans le chapitre précédent. Le schéma de la figure 4.1 résume l'approche retenue.

4.1 Choix de la stratégie d'insertion

La stratégie d'insertion obtenue par max-min est optimisée en fonction d'un couple $(D_{\mathsf{xy}}, D_{\mathsf{xy}'})$ donné. Or, dans les cas d'utilisation pratiques, on sait quelle distorsion d'insertion on peut tolérer en fonction de l'utilisation future de l'image marquée, mais on ne peut évaluer précisément la distorsion d'attaque. Le scénario le plus courant est que l'on souhaite transmettre un message de k bits et que l'on voudrait être le plus robuste possible pour une probabilité d'erreur maximale $\mathbf{P}_{\mathrm{e}}^{\max}$. Les courbes issues de nos expérimentations permettent de trouver le meilleur compromis. Les résultats présentés ici utilisent trois niveaux de compositions en ondelettes et le schéma prenant en compte l'information adjacente.

Supposons que nous voulions transmettre $k = 128$ bits dans l'image *Lena* avec une probabilité d'erreur par bit inférieure à $\mathbf{P}_{\mathrm{e}}^{\max} = 10^{-5}$. Nous nous autorisons une distor-

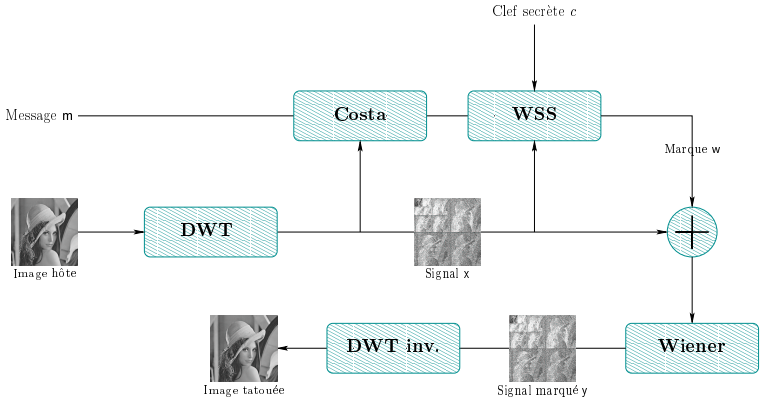


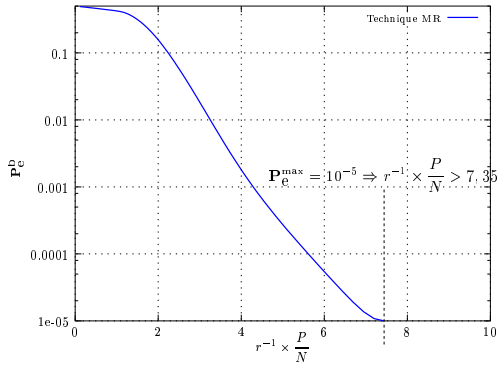
FIG. 4.1 – Insertion de la marque dans une image

sion d'insertion $D_{\mathsf{xy}} = 10$. Notre dictionnaire structuré est de rendement $r = k/n = 1/3$, et sa courbe de performance est donnée par la figure 4.2(a) (reprise de la figure 2.14(b) de la page 135). On voit que pour dépasser notre performance minimale imposée, il faut un rapport $r^{-1} \times P/N \geq 7,35$. Pour nos 128 bits, il faut donc $P/N \geq 7,35 \times 128 \simeq 951$. On reporte ce résultat sur la courbe de performance de notre stratégie d'insertion (courbe de la figure 4.2(b), reprise des résultats de la page 116) et on observe que la distorsion maximale à laquelle il est possible de résister est $D_{\mathsf{xy}'} \simeq 75$ (PSNR de $29,4$ dB). Il faudra donc utiliser la stratégie d'insertion adaptée à ce niveau de distorsion (en recherchant le couple (λ, χ) correspondant). Pour toute attaque de distorsion inférieure à 75, nous pouvons assurer la transmission de 128 bits utiles avec une probabilité d'erreur par bit de 10^{-5} (soit une probabilité d'erreur par message d'environ 10^{-3}).

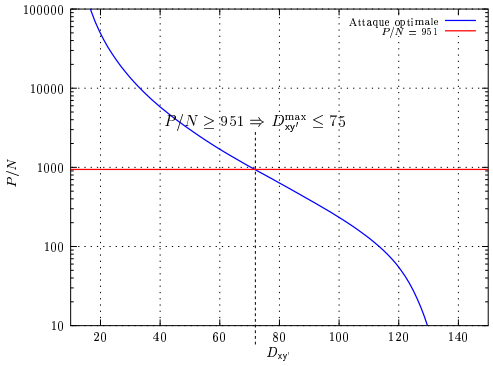
Pour certains traitements spécifiques, cette technique de recherche de stratégie peut être adaptée. Ainsi, si l'image tatouée est susceptible d'être recadrée (*cropping*), il est préférable de se laisser une marge. Dans l'exemple ci-dessus, il vaut mieux alors assurer un E_b/N_0 supérieur à 1500 (en supposant qu'un tiers de l'image soit supprimé par le recadrage). On choisira alors la stratégie d'insertion correspondant à $D_{\mathsf{xy}'} = 62$ (environ 30 dB).

4.2 Applications visées

Même s'il est conçu pour être le plus robuste possible, le fait que notre schéma soit optimisé pour une distorsion d'attaque donnée lui donne accès à plusieurs types d'applications : tatouage robuste classique, insertion d'informations supplémentaires pour enrichir le document hôte ou encore détection de marque.



(a) Valeur de P/N pour une probabilité d'erreur par bit de 10^{-5}



(b) Distorsion d'attaque maximale que peut supporter le schéma pour cette performance (image *Lena*)

FIG. 4.2 – Recherche de la distorsion d'attaque à viser pour une probabilité d'erreur donnée

4.2.1 Gestion de droits

L'application la plus évidente du tatouage robuste est la gestion de droits. Dans ce premier cas, la marque doit être très résistante afin que si d'éventuels pirates tentent de supprimer le message, l'image marquée soit tellement dégradée qu'elle ne présente plus de valeur. La taille du message est limitée. Il peut s'agir d'un nom ou d'un identifiant d'auteur, avec éventuellement quelques informations supplémentaires (date de création, bit indiquant si la copie est autorisée ou non, ...). Un système commercial d'authentification comme *Image bridge* [Dig] utilise une centaine de bits pour transmettre ces informations.

Cas général

Prenons $k = 128$ bits et une distorsion d'insertion de $D_{\mathsf{xy}} = 10$ (soit un PSNR supérieur à 38 dB). Notre schéma (avec prise en compte de l'information adjacente) nous assure d'extraire correctement la marque avec une probabilité d'erreur par bit égale à 10^{-5}, malgré une attaque de distorsion $D_{\mathsf{xy}'} = 75$ (PSNR de $29,4$ dB) pour l'image *Lena*, 30 pour *Rose* (PSNR de $33,5$ dB), 94 pour *Paper* ($28,4$ dB) et > 250 pour *Baboon* (< 24 dB). Toutes ces images sont de taille 512×512.

Avec désynchronisation

Nous conservons les paramètres $k = 128$ et $D_{\mathsf{xy}} = 10$ et utilisons la stratégie d'insertion prenant en compte une désynchronisation géométrique. S'il y a une erreur de recalage de $\Delta = 1$ pixel, nous pouvons tout de même résister à une attaque de distorsion $D_{\mathsf{xy}'} = 55$ (soit un PSNR de $30,7$ dB). Pour un décalage de 2 pixels, l'attaque maximale

est de distorsion 23 (34, 5 dB). Par contre, il nous est impossible d'assurer $\mathbf{P}_{\mathrm{e}}^{\max} = 10^{-5}$ pour $\Delta \geq 3$. Il faut alors augmenter le niveau de décomposition en ondelettes. Avec cinq niveaux et malgré un décalage de 5 pixels, on peut alors résister à une attaque de distorsion 15 (36, 4 dB).

4.2.2 Contenus enrichis

Enrichir des contenus est une application quasi opposée : une grande quantité d'information est transmise, mais la robustesse n'est pas capitale. Ce type de scénario est utilisé pour offrir à l'utilisateur du document un service supplémentaire. On peut penser par exemple à une adresse Internet au sein d'une image afin de visiter la galerie ou encore à des sous-titres dans une vidéo.

Même si la robustesse n'est pas importante, le document marqué peut subir quelques traitements (conversions de format, bruit lors de la transmission, ...). Prenons une robustesse à une distorsion d'environ 35 dB de PSNR ($D_{\mathsf{xy}'} = 20$). Avec une distorsion d'insertion $D_{\mathsf{xy}} = 5$ (PSNR supérieur à 40 dB), notre schéma nous permet de transmettre 345 octets ($k = 2760$) dans *Lena* ou encore 50 octets dans *Rose* (ce qui est suffisant pour une adresse Internet), avec une probabilité d'erreur par bit égale à 10^{-5}.

4.2.3 Détection de marque

Le point important dans la détection de marque est la probabilité de fausse alarme, c'est-à-dire la probabilité que l'on détecte que la marque est présente alors qu'elle ne l'est pas (voir la section 1.3 de la partie 1). La sécurité que l'on souhaite est réglée par la taille du message que l'on insère : pour un message de k bits, $\mathbf{P}_{\mathrm{f}} = 2^{-k}$. Avec une distorsion d'insertion $D_{\mathsf{xy}} = 10$ et un message de 25 bits, la probabilité de ne pas retrouver la marque si *Lena* est effectivement marquée est inférieure à $10^{-3,6}$ malgré une attaque de distorsion maximale $D_{\mathsf{xy}'} = 105$ (27, 9 dB). La probabilité de détecter la marque alors que l'image n'a pas été marquée est $\mathbf{P}_{\mathrm{f}} < 10^{-7}$.

Enfin, il est possible de mélanger extraction et détection. On peut alors transmettre un message et s'assurer que ce message est bien valide. Ainsi, si on souhaite transmettre 128 bits avec la même sécurité que pour le cas ci-dessus, il faut ajouter insérer $k = 128+25$ bits. Avec une distorsion d'insertion égale à 10, nous pouvons extraire la marque avec un probabilité d'erreur par bit égale à 10^{-5} malgré une attaque de distorsion 60 (PSNR de 30, 3 dB), et la probabilité de déclarer la marque valide alors que l'image n'a pas été marquée est $\mathbf{P}_{\mathrm{f}} < 10^{-7}$.

4.3 Impact visuel du marquage et de l'attaque

Les figures 4.3 à 4.6 présentent des exemples d'images marquées et attaquées. On remarque les différences de répartition suivant l'attaque visée. Pour les images des figures 4.3(a) et 4.5(a), la stratégie d'insertion est choisie afin de résister au mieux à une distorsion d'attaque de 20 et de respecter la distorsion d'insertion de 10 (38, 13 dB). Les images tatouées sont de bonne qualité : la marque est relativement bien étalée

(a) Image marquée avec la distorsion $D_{\mathsf{xy}} = 10$

(b) Détail de l'image marquée

(c) Image attaquée avec $D_{\mathsf{xy}'} = 20$

FIG. 4.3 – Insertion et attaque d'une marque sur *Lena* avec nos stratégies ($\varphi_i = 1$). La distorsion d'attaque visée est $D_{\mathsf{xy}'} = 20$ (PSNR de 35 dB)

avec une légère concentration sur les contours (forte variance du signal hôte). Par contre, lorsque l'on vise une distorsion d'attaque supérieure, l'énergie de la marque est essentiellement distribuée sur les échantillons de forte énergie. Comme nous utilisons le domaine DWT pour ces résultats, cela se manifeste sous la forme de rebonds autour des contours. Malgré une distorsion d'insertion identique, la marque est plus visible qu'avec la première stratégie.

Les images attaquées montrent bien les deux aspects de notre attaque optimale : ajout de bruit et filtrage. On voit que de nombreux détails disparaissent, même pour l'attaque la moins forte testée ici. Ainsi, le grain de la photo, très visible sur l'épaule de *Lena*, est totalement supprimé sur l'image 4.3(c), laissant une zone quasi uniforme. Pour les attaques les plus fortes (figures 4.4(c) et 4.6(c)), de nombreux échantillons sont annulés et l'image est rendue floue.

Les rebonds peuvent être atténués par l'introduction d'une pondération perceptuelle. Nous utilisons une mesure inspirée de celle de Watson, présentée dans la section 2.3.2 de la première partie. Elle est de la forme

$$\varphi_i^2 \propto \frac{1}{\sigma_{b_i}^2 + V_i^2} \tag{4.1}$$

$$\text{avec } V_i = \frac{1}{\|\Phi_i\|} \sum_{j \in \Phi_i} |x_j|^\rho. \tag{4.2}$$

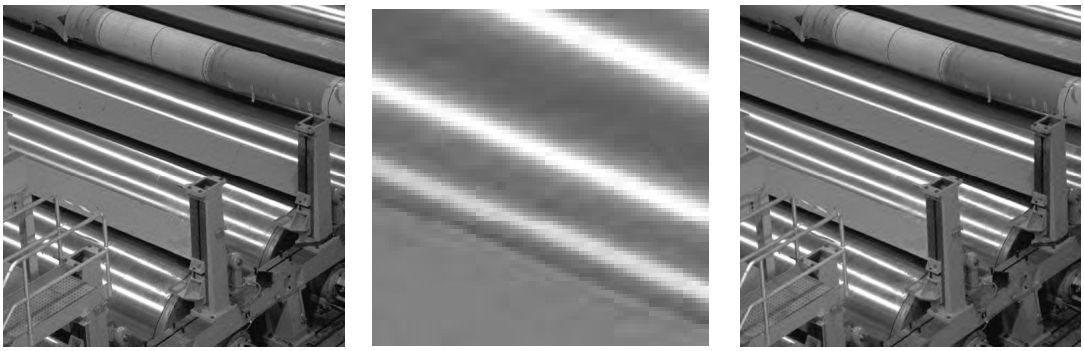
La formule de V_i ressemble à celle que nous utilisons pour calculer la variance des échantillons (calcul dans une fenêtre centrée sur l'échantillon considéré). On voit donc que les échantillons de variances les plus fortes auront une mesure φ_i faible. D'après la forme de l'insertion que nous utilisons, l'énergie de la marque sera moins concentrée

(a) Image marquée avec la distorsion $D_{\mathsf{xy}} = 10$

(b) Détail de l'image marquée

(c) Image attaquée avec $D_{\mathsf{xy}'} = 100$

FIG. 4.4 – Insertion et attaque d'une marque sur *Lena* avec nos stratégies ($\varphi_i = 1$). La distorsion d'attaque visée est $D_{\mathsf{xy}'} = 100$ (PSNR de 28 dB)

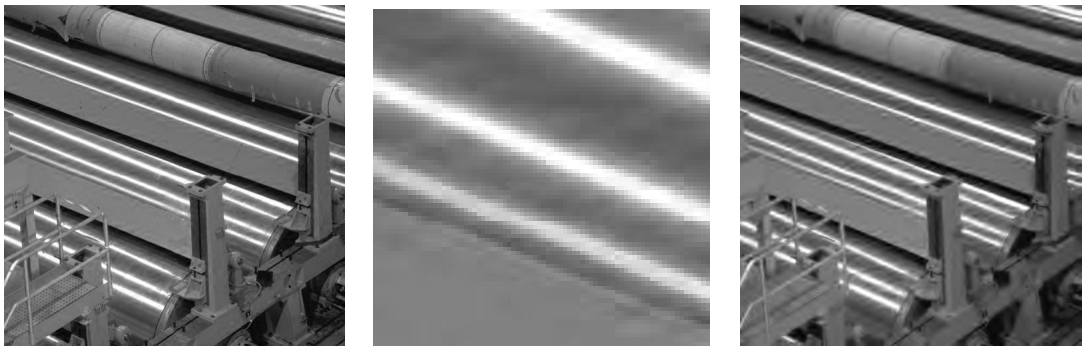


(a) Image marquée avec la distorsion $D_{\mathsf{xy}} = 10$

(b) Détail de l'image marquée

(c) Image attaquée avec $D_{\mathsf{xy}'} = 20$

FIG. 4.5 – Insertion et attaque d'une marque sur *Paper* avec nos stratégies ($\varphi_i = 1$). La distorsion d'attaque visée est $D_{\mathsf{xy}'} = 20$ (PSNR de 35 dB)



(a) Image marquée avec la distorsion $D_{\mathsf{xy}} = 10$

(b) Détail de l'image marquée

(c) Image attaquée avec $D_{\mathsf{xy}'} = 100$

FIG. 4.6 – Insertion et attaque d'une marque sur *Paper* avec nos stratégies ($\varphi_i = 1$). La distorsion d'attaque visée est $D_{\mathsf{xy}'} = 100$ (PSNR de 28 dB)

sur les basses fréquences. Cela est confirmé par les figures 4.7 et 4.8 : les rebonds sont moins visibles que pour les images des figures 4.4(a) et 4.6(a).

Conclusion

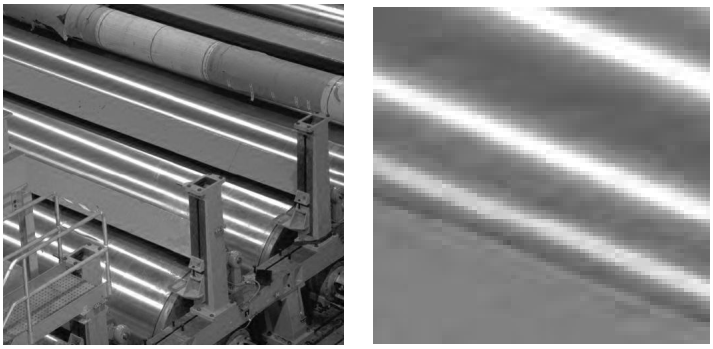
Nous avons montré comment utiliser les résultats des expérimentations des chapitres précédents pour choisir les meilleurs compromis entre probabilité d'erreur et force d'attaque supportable. Les résultats pratiques que nous avons exposés montrent que notre technique de tatouage est adaptée à plusieurs scénarii d'utilisation : gestion de droits, détection de marque, enrichissement de contenus, ... Le schéma présenté ici (tatouage d'images et utilisation de la transformée en ondelettes) est implémenté par le logiciel χ-mark[2], déposé à l'Agence de Protection des Programmes[1] et utilisé dans le projet RNRT Diphonet.

[1]Numéro IDDN.FR.001.480027.000.S.P.2002.000.41100.

(a) Image marquée avec la distorsion $D_{\mathsf{xy}} = 10$

(b) Détail de l'image marquée

FIG. 4.7 – Image *Lena* marquée avec notre stratégie (φ_i calculé par la mesure de Watson). La distorsion d'attaque visée est $D_{\mathsf{xy}'} = 100$ (PSNR de 28 dB)



(a) Image marquée avec la distorsion $D_{\mathsf{xy}} = 10$

(b) Détail de l'image marquée

FIG. 4.8 – Image *Paper* marquée avec notre stratégie (φ_i calculé par la mesure de Watson). La distorsion d'attaque visée est $D_{\mathsf{xy}'} = 100$ (PSNR de 28 dB)

Conclusion

Notre objectif était de définir un schéma de tatouage robuste s'appuyant sur des bases théoriques solides, et surtout de pouvoir en tirer une version pratique, facile à implémenter, dont les performances pourraient être proches des limites théoriques. Le tatouage a été abordé comme un problème de communication, et l'interaction en phase d'insertion et attaques a été modélisée par la théorie des jeux. Nous avons décomposé le problème en deux parties : la recherche d'une stratégie d'insertion optimale (comment répartir l'énergie de la marque sur le document hôte afin d'offrir les meilleures performances possibles), et la mise en œuvre pratique du schéma de Costa. Nous faisons ici la synthèse de nos contributions, puis donnons quelques perspectives d'évolution.

Synthèse de notre approche

Nous nous somme placés dans le cadre du tatouage robuste et aveugle. Plusieurs travaux ont étudié les limites de performance théorique de ce type de schéma. Néanmoins, les résultats proposés ne pouvaient être appliqués au sein d'un schéma pratique. Notre approche nous a permis de concevoir des solutions pratiques, guidées par des bases théoriques (codage canal et théorie des jeux), pouvant potentiellement atteindre les limites de performance définies par l'état de l'art.

Stratégie d'insertion

La stratégie d'insertion a été définie comme étant la répartition de l'énergie de la marque sur les échantillons du signal hôte. En utilisant une transmission par étalement de spectre et la théorie des jeux, nous avons proposé dans la seconde partie de ce manuscrit une stratégie d'insertion optimale adaptée aux signaux gaussiens non identiquement distribués. Pour cela,

- nous avons défini la stratégie d'attaque optimale, modélisée par un canal de type SAWGN (*scaling and additive white Gaussian noise*). Toutes les expérimentations montrent qu'elle est la plus performante de toutes les attaques testées (ajout de bruit, compression avec perte, filtrages, ...) : pour une distorsion donnée, cette attaque est celle qui réduit le plus les performances de la transmission,
- nous avons recherché la stratégie de défense (insertion de la marque) correspondante. Nos tests confirment le bien-fondé de cette approche et montrent qu'elle

est la stratégie la plus performante face à l'attaque optimale, apportant des gains très importants par rapport aux autres stratégies de l'état de l'art testées.

Pour une distorsion d'insertion et une distorsion d'attaque données, notre stratégie d'insertion définit une limite de performance de transmission minimale. Quelle que soit l'attaque appliquée sur le document marqué (et respectant la distorsion d'attaque donnée), il est par construction impossible de passer sous cette limite. Cela s'est vérifié dans nos expériences.

Codage pour canaux avec information adjacente

L'étalement de spectre que nous avons utilisé peut être vu comme la projection de signaux dans un sous-espace linéaire. Nous avons montré que cela définissait un canal gaussien avec information adjacente disponible à l'encodeur, propice à l'utilisation du schéma de Costa. Ce dernier définit une nouvelle mesure de performance que nous introduisons dans la résolution par théorie des jeux de la seconde partie. Reste que le schéma de Costa s'appuie sur un dictionnaire particulier, et ne peut être implémenté directement.

Nous avons donc vu dans la troisième partie de cette étude comment mettre en pratique le schéma de Costa, et notamment la construction du dictionnaire adéquat. En structurant un dictionnaire par l'introduction de bits d'index, nous avons proposé une technique innovante basée sur des codes correcteurs poinçonnés. Ce type de construction permet de s'adapter facilement à tout type de signal hôte en faisant varier le rendement du code. Les performances de notre codage s'approchent de celles de l'idéal de Costa (l'influence de l'information adjacente sur les performances est très réduite) et surpassent les codes correcteurs traditionnels.

De plus, nous avons introduit une nouvelle technique de suppression de l'interférence inter-symboles. Ce bruit additif introduit par l'étalement de spectre est pris en compte dans l'information adjacente grâce à un algorithme d'estimation itératif. Le gain en terme de performance est très sensible dans le cas d'attaques relativement faibles.

Extensions

La fin de l'étude s'est concentrée sur l'amélioration de nos stratégies d'insertion et d'attaques. Afin de résister aux attaques géométriques (désynchronisation du signal marqué), les schémas de tatouage utilisent un module de recalage. Or, celui-ci ne peut prévoir toutes les transformations possibles. Il en résulte des erreurs de synchronisation résiduelles. En modélisant ces erreurs, nous avons défini une nouvelle mesure de performance. Son introduction au sein de notre optimisation par théorie des jeux a abouti à une stratégie d'insertion inédite. Son application sur des signaux issus d'une transformée en ondelettes nous permet de trouver la répartition optimale de la marque entre les différentes sous-bandes de résolution. Ainsi, nous démontrons formellement qu'il est plus judicieux de marquer essentiellement les plus basses fréquences.

Nous apportons une autre amélioration en développant une attaque informée. À l'instar de l'insertion avec prise en compte de l'information adjacente, cette attaque uti-

lise la réalisation du signal à attaquer pour optimiser sa stratégie. Les expérimentations nous donnent des performances supérieures au cas non informé.

L'étalement de spectre associé à la théorie des jeux définit un canal gaussien au rapport signal-à-bruit maximal. Nous utilisons sur ce canal avec information adjacente le code structuré que nous avons développé. Grâce à cette combinaison, nous avons un schéma de tatouage complet dont la robustesse est garantie. En l'appliquant sur des signaux issus de la transformée en ondelettes d'image, nous avons implémenté un logiciel actuellement utilisé dans le projet RNRT Diphonet.

Perspectives

Nous avons introduit dans notre optimisation par max-min un paramètre perceptuel φ_i. Or il apparaît que la stratégie d'insertion optimale est croissante en fonction de celui-ci : plus l'échantillon marqué est perceptuellement important et plus il sera modifié par la marque. Bien sûr, l'insertion respecte au final notre contrainte de distorsion moyenne, mais nous avons remarqué dans les images marquées par cette technique des distorsions localisées assez visibles. La contrainte de distorsion moyenne n'est pas suffisante car l'énergie peut se concentrer sur quelques zones précises du document. Une amélioration sensible de la qualité pourrait être obtenue par la prise en compte d'un paramètre de type JND (seuil à partir duquel la distorsion devient perceptible), afin de limiter la distorsion maximale échantillon par échantillon.

La technique de construction de dictionnaire structuré que nous avons développée dans la troisième partie s'appuie sur un code convolutif. Nous avons montré que les performances atteintes par notre technique sont proches de celles qu'atteindrait ce code sans le bruit de l'information adjacente. Néanmoins, nous sommes relativement loin des limites de Shannon. Un gain important pourrait être obtenu rapidement en utilisant un code plus performant, tel qu'un turbo-code[2].

La façon de transmettre le paramètre i, qui détermine la taille des sous-dictionnaires, est restée en suspens. La solution la plus intéressante serait de le transmettre au sein même du mot de code. Nous pourrions ajouter en tête de notre motif d'*a priori* quelques bits (les premiers tests nous montrent que trois ou quatre bits seraient suffisants) correspondant à une version quantifiée de i. Ils seraient alors décodés en premier et le reste du treillis serait adapté en conséquence. L'influence de cette technique sur les performances du code n'a pas été précisement étudiée. Elle constitue une perspective de recherche prioritaire.

La résistance aux attaques géométriques est un aspect qui n'a pas été complètement étudié dans ces travaux. Nous avons certes développé une technique qui s'adapte aux erreurs de synchronisation légères, mais elle ne peut résister à de fortes transformations.

[2]Le passage à la version turbo montre un bond en avant important dans les travaux sur la construction de dictionnaires structurés de Chou *et al.* [CPR01].

La conception ou l'adaptation d'un outil de recalage nous donnerait accès à une gamme d'attaques beaucoup plus large.

Nous avons montré que l'attaque informée (prenant en compte la réalisation du signal reçu par l'attaquant) est plus performante que la version non-informée. Néanmoins, notre tentative pour trouver une défense appropriée n'a pas été particulièrement satisfaisante. Nous sommes passés par une approximation de la réalisation y des données marqués, et nous n'avons pu exhiber de solution analytique. La difficulté tient au fait que la réalisation de la marque w^{st} est dépendante des paramètres d'insertion que l'on cherche à optimiser. On retrouve une problématique relativement similaire à celle de l'estimation de l'interférence inter-symboles. Une voie à explorer pourrait donc être la résolution itérative du problème. L'initialisation se ferait en utilisant les paramètres $(\bar{\gamma}_i^\star, \sigma_{W_i}^\star)$ vus dans la section 3.2, et l'estimation de w^{st} (et donc de y) serait raffinée au fur et à mesure des itérations.

Annexe A

Développements de calculs

A.1 Calcul de l'estimateur optimal de la deuxième partie

Dans la deuxième partie, nous avons obtenu l'estimateur optimal selon le maximun *a posteriori* des bits insérés. L'estimation du $j^{\text{ème}}$ bit est définie par

$$\widehat{b_j} = \arg\max_b \{P_j(b)\} \tag{A.1}$$

$$\text{avec } P_j(b) = \Pr\left(B_j = b \,|\, \mathsf{Y}' = \mathsf{y}'\right). \tag{A.2}$$

Maximiser $P_j(b)$ correspond à maximiser un produit de probabilités défini par l'équation 1.14 de la page 66, et donc à minimiser $\Lambda(b)$:

$$\Lambda(b) = \sum_{i=1}^{m} \frac{\left(y'_i - \frac{\gamma_i \sigma_{W_i} \times \mathsf{G}(i,j) \times b}{\sqrt{n}}\right)^2}{\sigma_i^2}. \tag{A.3}$$

Le développement de $\Lambda(b)$ donne

$$\begin{aligned}
\Lambda(b) &= \sum_{i=1}^{m} \frac{y'^2_i}{\sigma_i^2} + \sum_{i=1}^{m} \frac{\gamma_i^2 \sigma_{W_i}^2 \times \mathsf{G}(i,j)^2 \times b^2}{n \times \sigma_i^2} \\
&\quad -2 \sum_{i=1}^{m} \frac{\gamma_i \sigma_{W_i} \times \mathsf{G}(i,j) \times b \times y'_i}{\sqrt{n} \times \sigma_i^2} \qquad (A.4) \\
&= b^2 \sum_{i=1}^{m} \frac{\gamma_i^2 \sigma_{W_i}^2 \times \mathsf{G}(i,j)^2}{n \times \sigma_i^2} - 2b \sum_{i=1}^{m} \frac{\gamma_i \sigma_{W_i} \times \mathsf{G}(i,j) \times y'_i}{\sqrt{n} \times \sigma_i^2} \\
&\quad + \left[\sum_{i=1}^{m} \frac{\gamma_i \sigma_{W_i} \times \mathsf{G}(i,j) \times y'_i}{\sqrt{n} \times \sigma_i^2}\right]^2 \\
&\quad + \sum_{i=1}^{m} \frac{y'^2_i}{\sigma_i^2} - \left[\sum_{i=1}^{m} \frac{\gamma_i \sigma_{W_i} \times \mathsf{G}(i,j) \times y'_i}{\sqrt{n} \times \sigma_i^2}\right]^2. \qquad (A.5)
\end{aligned}$$

En posant

$$\overline{b_j} = \frac{\sum_{i=1}^{m} \frac{\gamma_i \sigma_{W_i} \times \mathsf{G}(i,j) \times y'_i}{\sqrt{n} \times \sigma_i^2}}{\sum_{i=1}^{m} \frac{\gamma_i^2 \sigma_{W_i}^2}{n \times \sigma_i^2}} \quad \text{et} \quad \sigma_{b_j}^{-2} = \sum_{i=1}^{m} \frac{\gamma_i^2 \sigma_{W_i}^2}{n \times \sigma_i^2}, \tag{A.6}$$

on peut écrire l'équation A.5 sous la forme

$$\Lambda(b) = \frac{\left(b - \overline{b_j}\right)^2}{\sigma_{b_j}^2} + \Gamma. \tag{A.7}$$

La valeur de Γ est donnée par

$$\Gamma = \sum_{i=1}^{m} \frac{y_i'^2}{\sigma_i^2} - \left[\sum_{i=1}^{m} \frac{\gamma_i \sigma_{W_i} \times \mathsf{G}(i,j) \times y'_i}{\sqrt{n} \times \sigma_i^2}\right]^2 \times \left[\sum_{i=1}^{m} \frac{\gamma_i^2 \sigma_{W_i}^2}{n \times \sigma_i^2}\right]^{-1}, \tag{A.8}$$

et est indépendante de la valeur de b. Elle n'intervient donc pas dans la maximisation de l'équation A.1. Nous avons alors

$$\begin{aligned} \widehat{b_j} &= \arg\max_b \{P_j(b)\} \\ &= \arg\min_b \left\{ \frac{\left(b - \overline{b_j}\right)^2}{\sigma_{b_j}^2} \right\} \\ &= \overline{b_j}. \end{aligned} \tag{A.9}$$

A.2 Assignation des attaques aux domaines

La minimisation de la performance de la transmission aboutit à plusieurs stratégies définies des couples de paramètres (γ_i, σ_{Z_i}). Néanmoins, leur domaine d'application est parfois borné du fait de la contrainte $\sigma_{Z_i} > 0$. Cette section permet d'attribuer les stratégies aux domaines dans lesquelles elles sont le plus adaptées.

A.2.1 Sans prise en compte de l'information adjacente

La section 2.3.2 de la partie 2 montre que la solution trouvée par annulation des dérivées de la fonctionnelle J_λ^i (stratégie $\mathsf{a_I}$) n'est pas applicable sur toutes les données marquées. Cette solution n'est valide que sur un domaine, noté $\mathcal{D}_2$. Les contraintes de validité (équations 2.17 et 2.18 de la page 76) définissent deux autres domaines, notés $\mathcal{D}_1$ et $\mathcal{D}_3$. En examinant les cas limites parmi les valeurs possibles, on trouve en plus de la solution initiale deux autres stratégies d'attaques possibles. Nous avons au final :

- l'annulation du site, avec $\gamma_i = \sigma_{Z_i} = 0$, aboutissant à une valeur de fonctionnelle notée J_E (stratégie $\mathsf{a_E}$),
- l'attaque obtenue par annulation des dérivées, appelée stratégie intermédiaire, mélange de filtrage et d'ajout de bruit gaussien, donnant une valeur de fonctionnelle notée J_I (stratégie $\mathsf{a_I}$),

– et enfin le filtrage de Wiener ($\gamma_i = \gamma_i^{\mathrm{W}}$ et pas d'ajout de bruit), donnant J_W (stratégie $\mathsf{a_W}$).

Les expressions des fonctionnelles à minimiser pour ces trois attaques sont :

$$\begin{aligned} J_E &= \lambda\varphi_i^2\sigma_{X_i}^2 && \text{(A.10)} \\ J_I &= \gamma_i^{\mathrm{a}}\sqrt{\lambda}\varphi_i\sigma_{W_i} + \lambda\varphi_i^2\sigma_{X_i}^2\left(1-\gamma_i^{\mathrm{a}}\right) \\ &= J_E + \gamma_i^{\mathrm{a}}\sqrt{\lambda}\varphi_i\left(\sigma_{W_i} - \sqrt{\lambda}\varphi_i\sigma_{X_i}^2\right) && \text{(A.11)} \\ J_W &= \sqrt{\lambda}\varphi_i\frac{\sigma_{X_i}^2\sigma_{W_i}}{\sigma_{X_i}^2+\sigma_{W_i}^2} + \lambda\varphi_i^2\frac{\sigma_{X_i}^2\sigma_{W_i}^2}{\sigma_{X_i}^2+\sigma_{W_i}^2}. && \text{(A.12)} \end{aligned}$$

La formule de J_W peut se mettre sous la forme

$$\begin{aligned} J_W &= \sqrt{\lambda}\varphi_i\frac{\sigma_{X_i}^2\sigma_{W_i}}{\sigma_{X_i}^2+\sigma_{W_i}^2} + \lambda\varphi_i^2\sigma_{X_i}^2\left(1-\frac{\sigma_{X_i}^2}{\sigma_{X_i}^2+\sigma_{W_i}^2}\right) \\ &= J_E + \sqrt{\lambda}\varphi_i\frac{\sigma_{X_i}^2}{\sigma_{X_i}^2+\sigma_{W_i}^2}\left(\sigma_{W_i} - \sqrt{\lambda}\varphi_i\sigma_{X_i}^2\right). && \text{(A.13)} \end{aligned}$$

Et celle de J_I peut s'écrire

$$\begin{aligned} J_I &= J_E + \gamma_i^{\mathrm{a}}\sqrt{\lambda}\varphi_i\left(\sigma_{W_i} - \sqrt{\lambda}\varphi_i\sigma_{X_i}^2\right) \\ &= J_W - \sqrt{\lambda}\varphi_i\frac{\sigma_{X_i}^2}{\sigma_{X_i}^2+\sigma_{W_i}^2}\left(\sigma_{W_i} - \sqrt{\lambda}\sigma_{X_i}^2\right) + \gamma_i^{\mathrm{a}}\sqrt{\lambda}\varphi_i\left(\sigma_{W_i} - \sqrt{\lambda}\sigma_{X_i}^2\right) \\ &= J_W + \sqrt{\lambda}\varphi_i\left(\sigma_{W_i} - \sqrt{\lambda}\varphi_i\sigma_{X_i}^2\right)\left(\gamma_i^{\mathrm{a}} - \frac{\sigma_{X_i}^2}{\sigma_{X_i}^2+\sigma_{W_i}^2}\right). && \text{(A.14)} \end{aligned}$$

Sur le domaine $\mathcal{D}_1$, les seules solutions possibles sont l'annulation ou le filtrage de Wiener. Mais comme par définition $\sigma_{W_i} > \sqrt{\lambda}\varphi_i\sigma_{X_i}^2$ sur ce domaine, l'équation A.13 nous indique que $J_E < J_W$. L'annulation est donc le meilleur candidat.

Les trois attaques sont possibles sur $\mathcal{D}_2$. Or, comme sur ce domaine $\sigma_{W_i} \leq \sqrt{\lambda}\varphi_i\sigma_{X_i}^2$, J_I est inférieur ou égal à J_E depuis l'équation A.11. De plus, l'équation A.14 montre que $J_I \leq J_W$ car $\gamma_i^{\mathrm{a}} \leq \gamma_i^{\mathrm{w}}$. L'attaque intermédiaire est donc le meilleur candidat sur $\mathcal{D}_2$.

Enfin, seuls l'annulation et le filtrage de Wiener sont valides sur $\mathcal{D}_3$. On peut voir que $J_E \geq J_W$ sur ce domaine. L'attaque par filtrage de Wiener est le meilleur choix. La figure A.1 résume les associations entre domaines et attaques.

A.2.2 Avec prise en compte de l'information adjacente

L'étude de la fonctionnelle J_λ^i présentée dans la section 1.3.1 de la partie 3 nous a permis d'exhiber deux stratégies d'attaque. La première est l'attaque par annulation ($\bar{\bar{\gamma}}_i = 0$), donnant la fonctionnelle correspondante J_E. L'autre est l'attaque intermédiaire, mélange de bruit gaussien et de facteur d'échelle, dont les paramètres sont

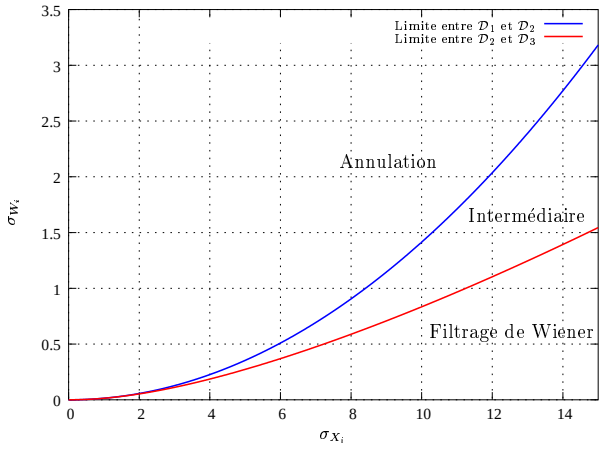


FIG. A.1 – Les trois domaines et les attaques associées dans l'optimisation de la seconde partie ($\lambda = 0,02$, $n = 1$ et $\varphi_i = 1$)

donnés par les équations 1.20 et 1.21 (page 109). La fonctionnelle correspondante est notée J_I. Les paramètres de ces deux attaques introduits dans l'expression de J_λ^i nous donne

$$J_E = \lambda\varphi_i^2\sigma_{X_i}^2 \tag{A.15}$$

$$\begin{aligned} J_I &= \gamma_i^{\mathrm{a}}\sqrt{\lambda}\varphi_i\sigma_{W_i} + \lambda\varphi_i^2\sigma_{X_i}^2\left(1-\gamma_i^{\mathrm{a}}\right) \\ &= J_E + \gamma_i^{\mathrm{a}}\sqrt{\lambda}\varphi_i\left(\sigma_{W_i} - \sqrt{\lambda}\varphi_i\sigma_{X_i}^2\right) \end{aligned} \tag{A.16}$$

Le fait que $\sigma_{Z_i} \geq 0$ définit deux domaines : le domaine $\mathcal{D}_2$ respectant la contrainte de l'équation 1.22 (page 109), et le domaine $\mathcal{D}_1$ ne la respectant pas. Sur le domaine $\mathcal{D}_1$, comme $\sigma_{W_i} > \sqrt{\lambda}\varphi_i\sigma_{W_i}^2$, l'équation A.16 nous indique que $J_E < J_I$. L'annulation est donc la meilleure attaque sur $\mathcal{D}_2$. Et pour le domaine $\mathcal{D}_1$, cette même équation montre que l'attaque intermédiaire est le meilleur choix. Cela est résumé par la figure A.2.

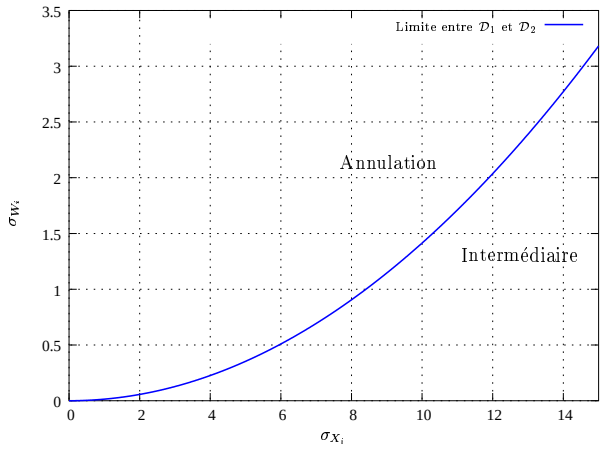


FIG. A.2 – Les trois domaines et les attaques associées dans l'optimisation de la troisième partie ($\lambda = 0,02$, $n = 1$ et $\varphi_i = 1$)

Annexe B

Interprétation géométrique du tatouage avec information adjacente

Une analyse et une interprétation géométrique complète de la communication avec information adjacente est donnée par Su *et al.* [SEG00]. Cette annexe en reprend une partie.

La figure B.1 donne une interprétation géométrique du schéma de Costa (ICS). Elle est présentée en deux dimensions mais doit être vue dans un espace n-dimensionnel. Le signal hôte x, qui représente l'information adjacente, est sur une hyper-sphère notée $\mathcal{S}_X$, centrée en $\mathsf{0}$ et de rayon $\sqrt{Q}$. Le dictionnaire $\mathcal{U}$ est constitué de mots de code d'énergie $P + \alpha^2 Q$ avec $\alpha = P/(P+N)$. Il est structuré en 2^{nC} sous-dictionnaires de taille $2^{nI(U;X)}$ (sur notre figure, nous avons trois sous-dictionnaires : carré, rond et triangle). Ses mots de codes sont disposés sur l'hyper-sphère $\mathcal{S}_{U/\alpha}$. À chaque mot de code est associé l'ensemble des points qui lui sont les plus proches. Cela définit des hyper-cônes de robustesse.

Dans notre exemple, nous cherchons à transmettre le message carré. Lors de l'insertion, le mot de code $\mathsf{u}^\star \in \mathcal{U}_\blacksquare$ le plus proche de x est recherché et le signal marqué est $\mathsf{y} = \mathsf{x} + \alpha(\mathsf{u}^\star/\alpha - \mathsf{x})$. L'insertion correspond à une homothétie de centre $\mathsf{u}^\star/\alpha$ et de rapport $1-\alpha$. Comme le montre la figure, cette technique permet de diriger n'importe quel x à l'intérieur du bon cône de robustesse. Dans le pire cas, x se trouve à équidistance de deux mots de $\mathcal{U}_\blacksquare$. Cela définit les arcs en gras de $\mathcal{S}_Y$ qui correspondent aux zones que peut atteindre y.

Notons le signal marqué $\mathsf{y} = \beta\mathsf{u}^\star + \mathsf{v}$, avec v bruit gaussien indépendant de U. En projetant y sur le mot de code $\mathsf{u}^\star$, on trouve

$$\begin{aligned} \beta &= \frac{\mathbb{E}\left[\langle \mathsf{y}, \mathsf{u}^\star \rangle\right]}{\|\mathsf{u}^\star\|^2} \\ &= \frac{\mathbb{E}\left[\langle \mathsf{w} + \mathsf{x}, \mathsf{w} + \alpha\mathsf{x} \rangle\right]}{\|\mathsf{u}^\star\|^2} \end{aligned}$$

$$= \frac{P + \alpha Q}{P + \alpha^2 Q}. \tag{B.1}$$

Et comme $\mathsf{v} = \mathsf{x} + \mathsf{w} - \beta \mathsf{u}^\star$, on montre facilement que

$$\sigma_V^2 = \frac{(1-\alpha)^2 PQ}{P + \alpha^2 Q}. \tag{B.2}$$

En utilisation la valeur $\alpha = P/(P+N)$, nous avons donc

$$\mathbb{E}\left[(\beta U)^2\right] = P\frac{(P+Q+N)^2}{(P+N)^2 + PQ} \tag{B.3}$$

$$\sigma_V^2 = N\frac{NQ}{(P+N)^2 + PQ}. \tag{B.4}$$

Selon le théorème du *sphere packing* [Lee67], lorsque n tend vers l'infini, nous pouvons disposer $2^{\frac{n}{2}\log_2(1+P/N)}$ sphères de rayon N ne se recouvrant pas sur une hyper-sphère de rayon P. Et comme

$$I(U;Y) = \frac{1}{2}\log_2\left[1 + \frac{P\,(P+Q+N)}{N\,(P+N)}\right] \tag{B.5}$$

avec

$$\frac{P\,(P+Q+N)}{N\,(P+N)} = \frac{P\frac{(P+Q+N)^2}{(P+N)^2+PQ}}{N\frac{NQ}{(P+N)^2+PQ} + N} = \frac{\mathbb{E}\left[(\beta U^\star)^2\right]}{\sigma_V^2 + N}, \tag{B.6}$$

nous avons $2^{nI(U;Y)}$ sphères de rayon au carré $\sigma_V^2 + N$ ne se recouvrant pas centrées sur une hyper-sphère de rayon au carré $\mathbb{E}\left[(\beta U)^2\right]$. Le signal marqué, déjà bruité par v, peut donc subir une attaque d'énergie N sans sortir statistiquement de l'hyper-cône de robustesse. Le schéma de Costa est donc similaire à la transmission du mot de code $\beta\mathsf{u}^\star$ soumis à deux bruits additifs : le bruit v introduit par le signal hôte et le bruit d'attaque z. Le rapport signal-à-bruit observé en sortie du canal correspond à celui de l'équation B.6, même si la capacité atteignable est équivalente à celle d'un canal gaussien classique de rapport P/N.

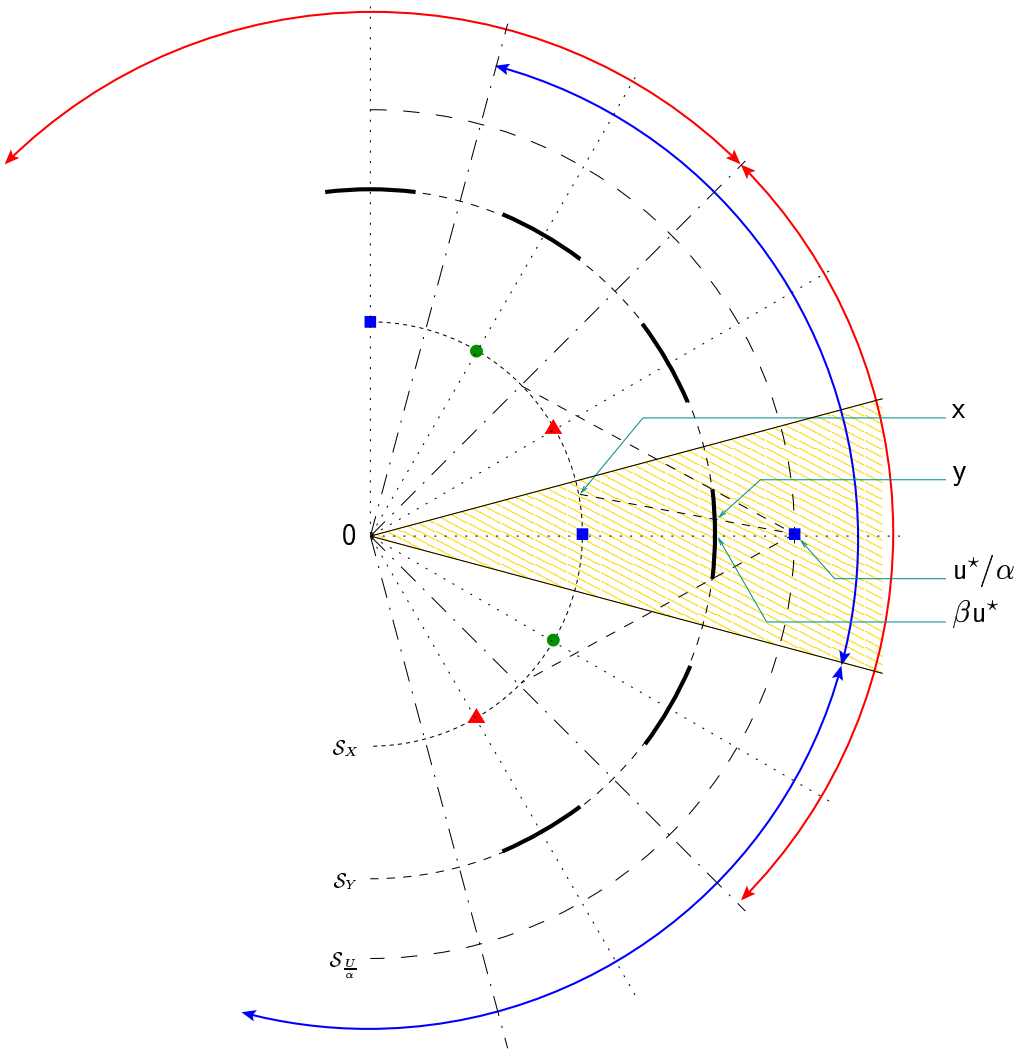


FIG. B.1 – Interprétation géométrique du schéma de Costa

Table des figures

Bibliographie

[Ade] E. H. Adelson. Lightness demonstrations. Illusions d'optique animées à voir sur `http ://www-bcs.mit.edu/gaz/`.

[Ade00] E. H. Adelson. *The new cognitive neurosciences*, chapter 24, pages 339–351. MIT Press, second edition, 2000.

[BBCP98] M. Barni, F. Bartolini, V. Cappellini, and A. Piva. A DCT-domain system for robust image watermarking. *IEEE Trans. Signal Proc. : Special Issue on Copyright Protection and Access Control for Multimedia Services*, 66(3) :357–372, 1998.

[BBRP99] M. Barni, F. Bartolini, A. De Rosa, and A. Piva. Capacity of the watermarking-channel : how many bits can be hidden within a digital image. In *Proc. SPIE*, volume 3657, pages 437–448, San Jose, CA, Jan. 1999.

[BCD98] P. Bas, J-M. Chassery, and F. Davoine. Using the fractal code to watermark images. In *Proc. Int. Conf. on Image Processing*, volume 1, pages 469–473, Chicago, IL, 1998.

[BDS+01] S. Baudry, J-F. Delaigle, B. Sankur, B. Macq, and H. Maitre. Analyses of error correction strategies for typical communication channels in watermarking. *IEEE Trans. Signal Proc.*, 81(6) :1239–1250, Jul. 2001.

[BGT93] C. Berrou, A. Glavieux, and P. Thitimajshima. Near shannon limit error-correcting coding and decoding : turbo codes. In *Proc. Int. Communications Conf.*, 1993.

[BNM02] S. Baudry, P. Nguyen, and H. Maître. Estimation of geometric distortions in digital watermarking. In *Proc. Int. Conf. on Image Processing*, Rochester, NY, Sep. 2002.

[CDRP99a] G. Csurka, F. Deguillaume, J. J. K. Ó Ruanaidh, and T. Pun. A bayesian approach to affine transformation resistant image and video watermarking. In *Proc. Int. Workshop on Information Hiding*, Dresden, Germany, Sep. 1999.

[CDRP99b] G. Csurka, F. Deguillaume, J. J. K. Ó Ruanaidh, and T. Pun. Tatouage d'images basé sur la transformée de Fourier discrète. In *Compression et représentation des signaux audiovisuels*, Sophia-Antipolis, France, Jun. 1999.

[CKLS97] I. J. Cox, J. Kilian, T. Leightom, and T. Shamoon. Secure spread spectrum watermarking for multimedia. *IEEE Trans. Image Proc.*, 6(12) :1673–1687, Dec. 1997.

[CL00] A. S. Cohen and A. Lapidoth. On the Gaussian watermarking game. In *Proc. Int. Symp. on Information Theory*, Sorrento, Italy, Jun. 2000.

[CL01] A. S. Cohen and A. Lapidoth. The capacity of the vector Gaussian watermarking game. In *Proc. Int. Symp. on Information Theory*, Washington, DC, Jun. 2001.

[CL02] A. S. Cohen and A. Lapidoth. The Gaussian watermarking game. *To appear in IEEE Trans. Info. Thy*, Jun. 2002.

[CMB02] I. J. Cox, M. L. Miller, and J. A. Bloom. *Digital watermarking*. Morgan Kaufmann Publishers, 2002.

[CNB98] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk. Wavelet-based statistical signal processing using hidden markov models. *IEEE Trans. Signal Proc. : Special Issue on Wavelets and Filterbanks*, Apr. 1998.

[Cos83] M. H. M. Costa. Writing on dirty paper. *IEEE Trans. Info. Thy*, 29(3) :439–441, May 1983.

[CPGR00] J. Chou, S. S. Pradhan, L. El Ghaoui, and K. Ramchandran. A robust optimization solution to the data hiding problem using distributed source coding principles. In *Proc. SPIE*, volume 3974, Jan. 2000.

[CPR99] J. Chou, S. S. Pradhan, and K. Ramchandran. On the duality between distributed source coding and data hiding. In *Proc. Conf. on Signals, System and Computers*, volume 2, pages 1503–1507, 1999.

[CPR01] J. Chou, S. S. Pradhan, and K. Ramchandran. Turbo coded treillis-based constructions for data embedding : channel coding with side information. In *Proc. Conf. on Signals, System and Computers*, Asilomar, CA, Nov. 2001.

[CT91] T. M. Cover and J. A. Thomas. *Elements of information theory*. Wiley-Interscience, Aug. 1991.

[CW00] B. Chen and G. W. Wornell. Preprocessed and postprocessed quantization index modulation methods for digital watermarking. In *Proc. SPIE*, San Jose, CA, Jan. 2000.

[CW01] B. Chen and G. W. Wornell. Quantization index modulation : a class of provably good methods for digital watermarking and information embedding. *IEEE Trans. Info. Thy*, 47(4) :1423–1443, May 2001.

[Del00] J-F. Delaigle. *Protection of intellectual property of images by perceptual watermarking*. PhD thesis, Université catholique de Louvain, Louvain-la-Neuve, Belgique, Sep. 2000.

[Des01] C. Desset. *Errors and losses in image transmission : error-correcting and recovering schemes*. PhD thesis, Université catholique de Louvain, Louvain-la-Neuve, Belgique, Jun. 2001.

[Dig] Digimarc. Image bridge. Site Internet de la compagnie : `http ://www.digimarc.com/`.

[Dix94] R. C. Dixon. *Spread spectrum systems with commercial applications.* Wiley-Interscience, Apr. 1994.

[DN00] I. Donescu and E. Nguyen. Combining visual and detection models in spread-spectrum watermarking. In *Proc. Int. Conf. on Image Processing*, 2000.

[EBG02] J. J. Eggers, R. Bäuml, and B. Girod. Estimation of amplitude modifications before SCS watermark detection. In *Proc. SPIE*, volume 4675, San Jose, CA, Jan. 2002.

[EBTG02] J. J. Eggers, R. Bäuml, R. Tzschoppe, and B. Girod. Scalar Costa scheme for information embedding. *IEEE Trans. Signal Proc. : Special Issue on Signal Processing for Data Hiding in Digital Media & Secure Content Delivery*, 2002.

[EG00] J. J. Eggers and B. Girod. Quantization watermarking. *Proc. SPIE*, Jan. 2000.

[EG02] J. J. Eggers and B. Girod. *Informed watermarking.* Kluwer Academic Publishers, 2002.

[EH83] A. El Gamal and C. Heegard. On the capacity of computer memories with defects. *IEEE Transactions on Information Theory*, 1983.

[ESG00] J. Eggers, J. Su, and B. Girod. Public key watermarking by eigenvectors of linear transforms. In *European Signal Proc. Conf.*, Tampere, Finland, Sep. 2000.

[FD03] T. Furon and P. Duhamel. An asymmetric watermarking method. *IEEE Trans. Signal Proc. : Special Issue*, 2003.

[FVD01] T. Furon, I. Venturi, and P. Duhamel. An unified approach of asymmetric schemes. In *Proc. SPIE*, San Jose, CA, 2001.

[Gal63] R. G. Gallager. *Low-Density Parity-Check Codes.* MIT Press, 1963.

[GN98] R. M. Gray and D. L. Neuhoff. Quantization. *IEEE Trans. Info. Thy*, 44(6), Oct. 1998.

[GP80] S. I. Gel'fand and M. S. Pinsker. Coding for channel with random parameters. *Problems of control and information theory*, 9(1) :19–31, 1980.

[GW92] R. C. Gonzales and R. E. Woods. *Digital image processing.* Addison-Wesley, 1992.

[hKRM99] M. K. Mıhçak, I. Kozintsev, K. Ranchandran, and P. Moulin. Low-complexity image denoising based on statistical modeling of wavelet coefficients. *IEEE Signal Proc. letters*, 6(12) :300–303, Dec. 1999.

[HVR01] A. Herrigel, S. Voloshynovskiy, and Y. Rytsar. The watermark template attack. In *Proc. SPIE*, San Jose, CA, Jan. 2001.

[HWS02] J. Huang, Y. Wang, and Y. Q. Shi. A blind audio watermarking algorithm with self-synchronisation. In *IEEE Int. Symp. on Circuits and Systems*, volume 3, pages 627–630, May 2002.

[JCL02] H-S. Jung, N-I. Cho, and S-U. Lee. Image adaptive watermarking based on warped discrete cosine transform. In *IEEE Int. Symp. on Circuits and Systems*, pages 209–212, May 2002.

[Ker83] A. Kerckoffs. La cryptographie militaire. *Journal des sciences militaires*, IX :5–38, Jan. 1883.

[KZ95] E. Koch and J. Zhao. Towards robust and hidden image copyright labeling. In *Proc. IEEE Workshop on Nonlinear signal and image processing*, pages 452–455, Halkidiki, Greece, Jun. 1995.

[LBC+00] C-Y. Lin, J. A. Blum, I. J. Cox, M. L. Miller, and Y. M. Lui. Rotation, scale, and translation resilient public watermarking for images. In *Proc. SPIE*, volume 3971, pages 90–98, San Jose, CA, 2000.

[LC97] C. Lin and S. Chang. A robust image authentication method distinguishing JPEG compression from malicious manipulation. Technical report, CU/CTR, Dec. 1997.

[LD99] E. Lin and E. Delp. A review of fragile image watermarks. In *Proc. of the Multimedia and Security Workshop*, pages 25–29, Orlando, FL, 1999.

[Lee67] J. Leech. Notes on sphere packings. *Canadian Journal of Mathematics*, 19 :251–267, 1967.

[LOJPG03] V. Licks, F. Ourique, R. Jordán, and F. Pérez-González. The effect of the random jitter attack on the bit error rate performance of spatial domain image watemarking. In *Proc. Int. Conf. on Image Processing*, Barcelona, Spain, Sep. 2003.

[Man01] A. Manoury. *Tatouage d'images numériques par paquets d'ondelettes*. PhD thesis, École Centrale, Nantes, France, Dec. 2001.

[MCB00] M. L. Miller, I. J. Cox, and J. A. Bloom. Informed embedding : exploiting image and detector information during watermark insertion. In *Proc. Int. Conf. on Image Processing*, Vancouver, Canada, Sep. 2000.

[MDC02] M. Miller, G. J. Doërr, and I. J. Cox. Dirty-paper trellis codes for watermarking. In *Proc. Int. Conf. on Image Processing*, Rochester, NY, Sep. 2002.

[MI01a] P. Moulin and A. Ivanović. Game-theoretic analysis of watermark dectection. In *Proc. Int. Conf. on Image Processing*, pages 975–978, Thessaloniki, Greece, Oct. 2001.

[MI01b] P. Moulin and A. Ivanović. The watermark selection game. In *Proc. Conf. on Info. Sciences and Systems*, Mar. 2001.

[MI03] P. Moulin and A. Ivanović. The zero-rate spread-spectrum watermarking game. *IEEE Trans. on Signal Processing*, 51(4) :1098–1117, Apr. 2003.

[MM01] P. Moulin and M. K. Mihçak. The data-hiding capacity of image sources. *preprint*, 2001.

[MM03] P. Moulin and M. K. Mihçak. A framework for evaluating the data-hiding capacity of image sources. *IEEE Trans. on Image Processing*, Mar. 2003.

[MO99] P. Moulin and J. A. O'Sullivan. Information-theoretic analysis of information hiding. *IEEE Trans. Info. Thy*, Oct. 1999.

[MO00] P. Moulin and J. A. O'Sullivan. Information-theoretic analysis of watermarking. Istanbul, Turkey, Jun. 2000.

[Mou01] P. Moulin. The parallel-Gaussian watermarking game. In *Proc. 35th Conf. on Information Sciences and Systems*, Baltimore, MD, Mar. 2001.

[MS74] J. L. Mannos and J. J. Sakrison. The effects of a visual fidelity criterion on the encoding of images. *IEEE Trans. Info. Thy*, IT-4 :525–536, 1974.

[Mül93] F. Müller. Distribution shape of two-dimensionnal DCT coefficient of natural images. *Electronic Letters*, (29) :1935–1936, Oct. 1993.

[OME98] J. A. O'Sullivan, P. Moulin, and J. M. Ettinger. Information-theoretic analysis of steganography. In *Proc. IEEE Symp. on Information Theory*, page 297, Cambridge, MA, Aug. 1998.

[Owe95] G. Owen. *Game theory.* Academic Press, third edition, 1995.

[PBBC97] A. Piva, M. Barni, F. Bartolini, and V. Cappellini. DCT-based watermark recovering without resorting to the uncorrupted original image. In *Proc. Int. Conf. on Image Processing*, volume 1, pages 520–523, Santa Barbara, CA, Oct. 1997.

[PBBC98] A. Piva, M. Barni, F. Bartonili, and V. Cappellini. Threshold selection for correlation-based watermark detection. In *Proc. COST 254 Workshop on Intelligent Communications*, pages 67–72, L'Aquila, Italy, Jun. 1998.

[Pet00] F. A. P. Petitcolas. Watermarking schemes evaluation. *IEEE Trans. Signal Proc.*, 17(5) :58–64, Sep. 2000.

[PGHB01] F. Pérez-Gonzáles, J. R. Hernández, and F. Balado. Approaching the capacity limit in image watermarking : a perspective on coding techniques for data hiding applications. *IEEE Trans. Signal Proc. : Special Issue on Information Theoretic Issues in Digital Watermarking*, 81(6), Jun. 2001.

[PJ96] J. Puate and F. Jordan. Using fractal compression scheme to embed a digital signature into an image. In *Proc. SPIE*, Boston, MA, 1996.

[PR00] S. S. Pradhan and K. Ramchandran. Distributed source coding using syndromes : design and construction. In *Proc. IEEE Data Compression Conf.*, pages 158–167, Mar. 2000.

[PRD+99] S. Pereira, J. J. K. Ó Ruanaidh, F. Deguillaume, G. Csurka, and T. Pun. Template based recovery of fourier-based watermarks using log-polar and log-log maps. In *IEEE Int. Conf. on Multimedia Computing and Systems*, Florence, Italy, Jun. 1999.

[PS98] H. C. Papadopulos and C-E. W. Sundberg. Simultaneous broadcasting of analog FM and digital audio signals by means of precancelling techniques. In *IEEE Int. Conf. on Communications*, volume 2, pages 728–732, 1998.

[PWM82] R. L. Pickholtz, J. M. Winograd, and L. B. Milstein. Theory of spread spectrum communications : a tutorial. *IEEE Trans. on Communications*, 30(5) :855–884, 1982.

[RBB+99] A. De Rosa, M. Barni, F. Bartolini, V. Cappellini, and A. Piva. Optimum decoding of non-additive full frame DFT watermarks. In *Proc. Int. Workshop on Information Hiding*, pages 160–172, Dresden, Germany, Sep. 1999.

[RDCD02] C. Rey, G. Doërr, G. Csurka, and J-L. Dugelay. Toward generic image dewatermarking ? In *Proc. Int. Conf. on Image Processing*, volume 3, pages 633–636, Rochester, NY, Sep. 2002.

[RP98] J. J. K. Ó Ruanaidh and T. Pun. Rotation, scale and translation invariant spread spectrum digital image watermarking. *IEEE Trans. Signal Proc.*, 66(3) :303–317, 1998.

[SC02] S. Somasundaram and R. Chandramouli. Perceptually based waterfilling for watermarking. In *IEEE Int. Symp. on Circuits and Systems*, volume 2, pages 456–459, May 2002.

[SEG00] J. K. Su, J. J. Eggers, and B. Girod. Channel coding and rate distortion with side information : geometric interpretation and illustration of duality. *Submitted to IEEE Trans. on Information Theory*, May 2000.

[SEG01a] J. K. Su, J. J. Eggers, and B. Girod. Analysis of digital watermarks subjected to optimum linear filtering and additive noise. *IEEE Trans. Signal Proc. : Special Issue on Information Theoretic Issues in Digital Watermarking*, 81(6), Jun. 2001.

[SEG01b] J. K. Su, J. J. Eggers, and B. Girod. Optimum attack on digital watermarks and its defense. In *Proc. Conf. on Signals, System and Computers*, Asilomar, CA, Oct. 2001.

[SG99] J. K. Su and B. Girod. Power spectrum condition for energy-efficient watermarking. In *Proc. Int. Conf. on Image Processing*, Kobe, Japan, Oct. 1999.

[Sim98] G. Simmons. The history of subliminal channels. *IEEE Journal on Selected Areas in Communications*, 16(4), May 1998.

[SKH02] K. Su, D. Kundur, and D. Hatzinakos. A novel approach to collusion-resistant video watermarking. In *Proc. SPIE*, volume 4675, pages 491–502, San Jose, CA, Jan. 2002.

[SPR98] S. D. Servetto, C. I. Podilchuk, and K. Ramchandran. Capacity issues in digital image watermarking. In *Proc. Int. Conf. on Image Processing*, volume 1, pages 445–449, Chicago, IL, Oct. 1998.

[Sta99] S. Stahl. *A gentle introduction to game theory.* American Mathematical Society, 1999.

[sti] Stirmark version 4.0. Logiciel d'évaluation de schémas de tatouage, téléchargeable sur `http ://www.petitcolas.net/fabien/watermarking/stirmark/`.

[Tau00] D. Taubman. High performance scalable image compression with EBCOT. *IEEE Trans. Image Proc.*, 9(7) :1158–1170, Jul. 2000.

[TM01] D. S. Taubman and M. W. Marcellin. *JPEG 2000 : image compression fundamentals, standards, and practice.* Kluwer Academic Publishers, Nov. 2001.

[Tur89] L. F. Turner. Digital data security system. Patent IPN WO 89/08915, 1989.

[VDP00] S. Voloshynovskiy, F. Deguillaume, and T. Pun. Content adaptive watermarking based on a stochastic multiresolution image modeling. In *European Signal Proc. Conf.*, Tampere, Finland, Sep. 2000.

[VDP01] S. Voloshynovskiy, F. Deguillaume, and T. Pun. Multibit digital watermarking robust against local nonlinear geometrical distortions. In *Proc. Int. Conf. on Image Processing*, pages 999–1002, Thessaloniki, Greece, Oct. 2001.

[VDPP01] S. Voloshynovskiy, F. Deguillaume, S. Pereira, and T. Pun. Optimal adaptive diversity watermarking with channel state estimation. In *Proc. SPIE*, San Jose, CA, Jan. 2001.

[VHBP99] S. Voloshynovskiy, A. Herrigel, N. Baumgärtner, and T. Pun. A stochastic approach to content adaptive digital image watermarking. In *Proc. Int. Workshop on Information Hiding*, pages 212–236, Dresden, Germany, Sep. 1999.

[Vit67] A. J. Viterbi. Error bounds for convolutional codes and asymptotically optimum decoding algorithm. *IEEE Trans. Info. Thy*, 13 :260–269, Apr. 1967.

[Wal91] G. K. Wallace. The JPEG still picture compression standard. *Communications of the ACM*, 34(4), Apr. 1991.

[Wat93] A. B. Watson. DCT quantization matrices visually optimized for individual images. *Proc. SPIE*, 1913 :202–216, 1993.

[WBL02] Z. Wang, A. C. Bovik, and L. Lu. Why is image quality assessment so difficult ? In *IEEE Conf. on Acoustics, Speech and Signal Processing*, volume 4, pages 3313–3316, Orlando, FL, May 2002.

[WPD99] R. Wolfgang, C. Podilchuk, and E. J. Delp. Perceptual watermarks for digital images and video. *Proc. IEEE*, 87(7), Jul. 1999.

[WV99] C. Weidmann and M. Vetterli. Rate-distortion analysis of Spike processes. In *Proc. Data Compression Conf.*, Snowbird, UT, Mar. 1999.

[WYL02] M. Wu, H. Yu, and B. Lui. Data hiding in image and video : part-II Designs and applications. *IEEE Trans. Image Proc.*, Jan. 2002.

[WYSV97] A. B. Watson, G. Y. Yang, J. A. Solomon, and J. Villasenor. Visibility of wavelet quantization noise. *IEEE Trans. Image Proc.*, 6(8) :1164–1175, 1997.

[ZL02] Y. Zhao and R. L. Lagendijk. Video watermarking scheme resistant to geometric attacks. In *Proc. Int. Conf. on Image Processing*, pages 145–149, Rochester, NY, Sep. 2002.

Publications

Journaux :

1. S. Pateux et G. Le Guelvouit, **Practical watermarking scheme based on wide spread spectrum and game theory**, *Signal Processing : Image Communication*, 18 : 283–296, Apr. 2003.

Congrès internationaux :

1. G. Le Guelvouit et S. Pateux, **Wide spread spectrum watermarking with side information and interference cancellation**, in *Proc. SPIE*, Santa Clara, CA, Jan. 2003.
2. G. Le Guelvouit, S. Pateux et C. Guillemot, **Perceptual watermarking of non i.i.d. signals based on wide spread spectrum using side information**, in *Proc. Int. Conf. on Image Processing*, Rochester, NY, Sep. 2002.
3. G. Le Guelvouit, S. Pateux et C. Guillemot, **Information-theoretic resolution of perceptual WSS watermarking of non i.i.d. Gaussian signals**, in *Proc. Eur. Signal Processing Conf.*, Toulouse, France, Sep. 2002.

Congrès nationaux :

1. G. Le Guelvouit, S. Pateux et J. Delhummeau, **Construction de codes pour tatouage avec prise en compte de l'information adjacente**, in *GRETSI Symp. on Image and Signal Processing*, Paris, France, Sep. 2003.
2. S. Pateux et G. Le Guelvouit, **Performances d'un système de tatouage soumis à des désynchronisations**, *GRETSI Symp. on Image and Signal Processing*, Paris, France, Sep. 2003.

Divers :

1. S. Pateux, G. Le Guelvouit et C. Guillemot, **Dispositif pour le marquage et la restitution de signaux multimedia**, brevet num. FR-0213605, Oct. 2002.

Abstract

Numerical technologies increase the ease of transmission, storage and modification of multimedia content, but also makes authentication and copyright management difficult. Robust watermarking is a solution that appeared about ten years ago. It embeds a message (author name, numerical signature, ...) within a host multimedia document. This process must not be perceptible in order not to spoil the regular use of the watermarked documents. Moreover, one must be able to correctly extract the message despite modifications of the host (*i.e.* attacks).

Although the first approaches were empirical, watermarking was quickly considered as a communication problem. This work deals with a novel watermarking scheme based on a wide spread spectrum (a communication technique that fits to very noisy channels) and considering SAWGN attacks (scaling and additive white Gaussian noise). The relationship between watermarking and attacks is modeled by a game between an attacker and a defender. A max-min optimization leads to the optimal attack and the corresponding embedding strategy (counter-attack). Experiments using signals from wavelet transformed images confirm the relevance of our approach : the attack is very efficient and the embedding strategy leads to better results than previous literature's approaches.

The watermarking channel is a channel with side information : a part of the noise (the host signal) is perfectly known during the embedding process. Previous works on this kind of channel showed better theoretical performances than on a classical Gaussian channels. In the third part of this work, we introduce a new performance measure in our watermarking game, taking into account the side information. We then develop a structured dictionary for this kind of channel. The experiments show great improvements in terms of performance. Finally, the last chapter deals with some game improvements : we study the influence of geometrical desynchronizations on the performance of our watermarking scheme, and we define an informed attack.

Keywords

Blind and robust watermarking, wide spread spectrum, SAWGN attacks, game theory, max-min optimization, channels with side information, structured dictionary, punctured codes, inter-symbols interference (ISI), geometrical desynchronization, wavelet transform.

Résumé

La technologie numérique rend la transmission, le stockage et la modification de documents multimedia beaucoup plus aisés qu'auparavant. Mais du fait de cette facilité, l'authentification et la gestion des droits d'auteur deviennent difficiles. Le tatouage robuste est une solution qui s'est beaucoup développée depuis une dizaine d'années. Il permet d'insérer au sein d'un document multimedia (document hôte) un message (identifiant d'auteur, signature numérique, ...). Cette modification est peu perceptible afin de ne pas gêner l'exploitation normale du document marqué. De plus, l'algorithme d'extraction doit pouvoir retrouver le message malgré d'éventuelles modifications de l'hôte (attaques).

Alors que les premières études sur ce domaine étaient empiriques, il est apparu rapidement que le tatouage était assimilable à un problème de communication. Nous faisons le choix d'utiliser une technique de transmission particulièrement adaptée aux canaux fortement bruités, appelée étalement de spectre, et nous considérons des attaques SAWGN (*scaling and additive white Gaussian noise*). L'interaction entre tatouage et attaques est modélisée par un jeu entre un attaquant et un défenseur. Une optimisation de type max-min donne la forme de l'attaque optimale, qui va le plus réduire la performance de la transmission, et de la stratégie qui permet de s'en protéger au mieux. L'application de ces résultats sur des signaux issus de la transformée en ondelettes d'images confirme le bien-fondé de notre approche : l'attaque obtenue est la plus efficace, et la stratégie d'insertion est bien plus performante que les techniques testées issues de l'état de l'art.

Le canal de tatouage est un canal avec information adjacente : une partie du bruit (le signal hôte) est connue au moment de l'insertion de la marque. Les travaux sur ces canaux promettent des performances théoriques bien supérieures aux canaux gaussiens classiques. Dans la troisième partie de ce manuscrit, nous reprenons l'optimisation par théorie des jeux en incluant une mesure de performance prenant en compte l'information adjacente. Nous détaillons ensuite la construction d'un dictionnaire structuré adapté à ces canaux. Les expérimentations nous indiquent de forts gains de performance. Enfin, le dernier chapitre améliore le jeu : l'étude de l'influence de désynchronisations géométriques sur les performances, et la définition d'une attaque informée prenant en compte la réalisation du signal reçu.

Mots clef

Tatouage aveugle robuste, étalement de spectre, attaques SAWGN, théorie du jeu, optimisation max-min, canaux avec information adjacente, dictionnaire structuré, codes poinçonnés, interférence inter-symboles (ISI), désynchronisation géométrique, transformée en ondelettes.